Query Match 100.0%; Score 20; DB 9; Length 2072;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 gcgtctctactgcctcttcg 20
           ||||||||||||||||||||
Db     162 GCGTCTCTACTGCCTCTTCG 181
```

SID 32

172 - 153 = SID 31

1315 - 1295 = SID 39

SID 52 also SID 89

RESULT 1
BC014484
LOCUS BC014484 1685 bp mRNA linear PRI 19-SEP-2001 not prior
DEFINITION Homo sapiens, Similar to dystonia 1, torsion (autosomal dominant; torsin A), clone MGC:23205 IMAGE:4869856, mRNA, complete cds.
ACCESSION BC014484
VERSION BC014484.1 GI:15680257
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1685)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 34 Row: l Column: 17
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES Location/Qualifiers
source 1. .1685
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:23205 IMAGE:4869856"
/tissue_type="Skin, melanotic melanoma, high MDR."
/clone_lib="NIH_MGC_49"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
CDS 20. .613
/codon_start=1
/product="Similar to dystonia 1, torsion (autosomal dominant; torsin A)"
/protein_id="AAH14484.1"

/db_xref="GI:15680258"
/translation="MKLGRAVLGLLLLAPSVVQAVEPISLGLALAGVLTGYIYPRLYC
LFAECCGQKRSLSREALQKDLDDNLFGQHLAKKIILNAVFGFINNPKPKKPLTLSLHG
WTGTGKNFVSKIIAENIYEGGLNSDYVHLFVATLHFPHASNITLYKARMEVWNPFLDV
IGFGVSLLWDEIWEFYVEMSEPGKRFMSQFPLERCRS"

BASE COUNT 379 a 421 c 425 g 460 t
ORIGIN

Query Match 98.0%; Score 392; DB 9; Length 1685;
Best Local Similarity 98.8%; Pred. No. 2.5e-117;
Matches 395; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 gaatatttacgagggtggtctgaacagtgactatgtccacctgtttgtggccacattgct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     370 GAATATTTACGAGGGTGGTCTGAACAGTGACTATGTCCACCTGTTTGTGGCCACATTGCA 429

Qy      61 ctttccacatgcttcaaacatcaccttgtacaaggcaaggatggaagtttggaatccctt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     430 CTTTCCACATGCTTCAAACATCACCTTGTACAAGGCAAGGATGGAAGTTTGGAATCCCTT 489

Qy     121 cctggatgtcatcgggtttggggtctctttgttgtgggatgagatttgggagttctatgt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     490 CCTGGATGTCATCGGGTTTGGGGTCTCTTTGTTGTGGGATGAGATTTGGGAGTTCTATGT 549

Qy     181 tgaaatgagtgagcccggaaaacggttcatgtctcagttccccttggaaaggtgtagaag 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     550 TGAAATGAGTGAGCCCGGAAAACGGTTCATGTCTCAGTTCCCCTTGGAAAGGTGTAGAAG 609

Qy     241 ttaagagtttgagatgcgtggagcagttaataccatcaaagctttgtggtgggttctgaa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     610 TTAAGAGTTTGAGATGCGTGGAGCAGTTAATACCATCAAAGCTTTGTGGTGGGTTCTGAA 669

Qy     301 aatcggtccagtgagtatgtagggtcatgggattttagaggtggacatgatcaaatccat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     670 AATCGGTCCAGTGAGTATGTAGGGTCATGGGATTTTAGAGGTGGACATGATCAAATCCAT 729

Qy     361 cttagagatcaacacatctcactcatttttatttttcttat 400  + + 402
          |||||||||||||||||||||||||||| | || || | |        |
Db     730 CTTAGAGATCAACACATCTCACTCATTTTTTTATTTTTTT 769  C + 771
```

SID 89

RESULT 2
AC027008
LOCUS AC027008 166889 bp DNA linear HTG 08-NOV-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-212N14 map 8, WORKING DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION AC027008
VERSION AC027008.4 GI:10280898
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166889)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-212N14

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166889)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 23, 2000 this sequence version replaced gi:8080819.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8771
                Center clone name: 212_N_14
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 150653 bases at least Q40
                Consensus quality: 158747 bases at least Q30
                Consensus quality: 161880 bases at least Q20
                Insert size: 186000; agarose-fp
                Insert size: 164389; sum-of-contigs
                Quality coverage: 3.6 in Q20 bases; agarose-fp
                Quality coverage: 4.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
            * be preserved.
            *        1      140: contig of 140 bp in length
            *      141 240: gap of     100 bp
            *      241     1566: contig of 1326 bp in length
            *     1567 1666: gap of      100 bp
            *     1667    26279: contig of 24613 bp in length
            *    26280 26379: gap of      100 bp
            *    26380    27676: contig of 1297 bp in length
            *    27677 27776: gap of      100 bp
            *    27777    29820: contig of 2044 bp in length
            *    29821 29920: gap of      100 bp
            *    29921    33216: contig of 3296 bp in length
            *    33217 33316: gap of      100 bp
            *    33317    36627: contig of 3311 bp in length
            *    36628 36727: gap of      100 bp
            *    36728    39382: contig of 2655 bp in length
            *    39383 39482: gap of      100 bp
            *    39483    42417: contig of 2935 bp in length
            *    42418 42517: gap of      100 bp
            *    42518    46306: contig of 3789 bp in length
            *    46307 46406: gap of      100 bp
            *    46407    50207: contig of 3801 bp in length
            *    50208 50307: gap of      100 bp
            *    50308    53363: contig of 3056 bp in length
            *    53364 53463: gap of      100 bp
            *    53464    56760: contig of 3297 bp in length
            *    56761 56860: gap of      100 bp
            *    56861    61207: contig of 4347 bp in length
            *    61208 61307: gap of      100 bp
            *    61308    65984: contig of 4677 bp in length
            *    65985 66084: gap of      100 bp
            *    66085    72072: contig of 5988 bp in length
            *    72073 72172: gap of      100 bp
            *    72173    77741: contig of 5569 bp in length
            *    77742 77841: gap of      100 bp
            *    77842    85850: contig of 8009 bp in length
            *    85851 85950: gap of      100 bp
            *    85951    92902: contig of 6952 bp in length
            *    92903 93002: gap of      100 bp
            *    93003   103668: contig of 10666 bp in length
            *   103669 103768: gap of      100 bp
            *   103769   109322: contig of 5554 bp in length
            *   109323 109422: gap of      100 bp
            *   109423   118526: contig of 9104 bp in length
            *   118527 118626: gap of      100 bp
            *   118627   128874: contig of 10248 bp in length
            *   128875 128974: gap of      100 bp
            *   128975   138016: contig of 9042 bp in length
            *   138017 138116: gap of      100 bp
            *   138117   166500: contig of 28384 bp in length
            *   166501 166600: gap of      100 bp
            *   166601   166889: contig of 289 bp in length.
FEATURES             Location/Qualifiers
     source          1. .166889
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
```

```
                     /map="8"
                     /clone="RP11-212N14"
                     /clone_lib="RPCI-11 Human Male BAC"
misc_feature         1. .140
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
misc_feature         241. .1566
                     /note="assembly_fragment"
misc_feature         1667. .26279
                     /note="assembly_fragment"
misc_feature         26380. .27676
                     /note="assembly_fragment"
misc_feature         27777. .29820
                     /note="assembly_fragment"
misc_feature         29921. .33216
                     /note="assembly_fragment"
misc_feature         33317. .36627
                     /note="assembly_fragment"
misc_feature         36728. .39382
                     /note="assembly_fragment"
misc_feature         39483. .42417
                     /note="assembly_fragment"
misc_feature         42518. .46306
                     /note="assembly_fragment"
misc_feature         46407. .50207
                     /note="assembly_fragment"
misc_feature         50308. .53363
                     /note="assembly_fragment"
misc_feature         53464. .56760
                     /note="assembly_fragment"
misc_feature         56861. .61207
                     /note="assembly_fragment"
misc_feature         61308. .65984
                     /note="assembly_fragment"
misc_feature         66085. .72072
                     /note="assembly_fragment"
misc_feature         72173. .77741
                     /note="assembly_fragment"
misc_feature         77842. .85850
                     /note="assembly_fragment"
misc_feature         85951. .92902
                     /note="assembly_fragment"
misc_feature         93003. .103668
                     /note="assembly_fragment"
misc_feature         103769. .109322
                     /note="assembly_fragment"
misc_feature         109423. .118526
                     /note="assembly_fragment"
misc_feature         118627. .128874
                     /note="assembly_fragment"
misc_feature         128975. .138016
                     /note="assembly_fragment"
misc_feature         138117. .166500
                     /note="assembly_fragment"
misc_feature         166601. .166889
                     /note="assembly_fragment
```

clone_end:T7
vector_side:right"
BASE COUNT 43782 a 38337 c 39037 g 43225 t 2508 others
ORIGIN

Query Match 98.0%; Score 392; DB 2; Length 166889;
Best Local Similarity 98.8%; Pred. No. 4.3e-117;
Matches 395; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 gaatatttacgagggtggtctgaacagtgactatgtccacctgtttgtggccacattgct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117201 GAATATTTACGAGGGTGGTCTGAACAGTGACTATGTCCACCTGTTTGTGGCCACATTGCA 117260

Qy      61 ctttccacatgcttcaaacatcaccttgtacaaggcaaggatggaagtttggaatccctt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117261 CTTTCCACATGCTTCAAACATCACCTTGTACAAGGCAAGGATGGAAGTTTGGAATCCCTT 117320

Qy     121 cctggatgtcatcgggtttggggtctctttgttgtgggatgagatttgggagttctatgt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117321 CCTGGATGTCATCGGGTTTGGGGTCTCTTTGTTGTGGGATGAGATTTGGGAGTTCTATGT 117380

Qy     181 tgaaatgagtgagcccggaaaacggttcatgtctcagttccccttggaaaggtgtagaag 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117381 TGAAATGAGTGAGCCCGGAAAACGGTTCATGTCTCAGTTCCCCTTGGAAAGGTGTAGAAG 117440

Qy     241 ttaagagtttgagatgcgtggagcagttaataccatcaaagctttgtggtgggttctgaa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117441 TTAAGAGTTTGAGATGCGTGGAGCAGTTAATACCATCAAAGCTTTGTGGTGGGTTCTGAA 117500

Qy     301 aatcggtccagtgagtatgtagggtcatgggattttagaggtggacatgatcaaatccat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117501 AATCGGTCCAGTGAGTATGTAGGGTCATGGGATTTTAGAGGTGGACATGATCAAATCCAT 117560

Qy     361 cttagagatcaacacatctcactcatttttatttttcttat 400
          |||||||||||||||||||||||||||||| | || || | |
Db  117561 CTTAGAGATCAACACATCTCACTCATTTTTTTATTTTTTT 117600
```

T T
C T

52,53,
SID 52 & 89

RESULT 3
AL158207/c
LOCUS AL158207 169963 bp DNA linear PRI 25-SEP-2001
DEFINITION Human DNA sequence from clone RP11-409K20 on chromosome 9 Contains the TOR1B gene for torsin family 1 member B (torsin B) (DQ1), the DYT1 gene for 'dystonia 1, torsion' (autosomal dominant; torsin A) (DQ2, TOR1A), the gene for hepatocellular carcinoma-associated antigen 59 (HSPC220, LOC51759), the USP20 gene for ubiquitin specific protease 20 (KIAA1003), and the gene for formin-binding protein 17 (FBP17, includes KIAA0554, FLJ13619, FLJ10754 and FLJ10113). Contains ESTs, STSs, GSSs and four CpG islands, complete sequence.
ACCESSION AL158207
VERSION AL158207.15 GI:12717949
KEYWORDS HTG; CpG island; DQ1; dystonia; DYT1; FBP17; FLJ10113; FLJ10754; FLJ13619; formin-binding; hepatocellular; HSPC220; KIAA0554; KIAA1003; LOC51759; protease; TOR1B; torsin; ubiquitin; USP20.
SOURCE human.

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169963)
  AUTHORS   Babbage,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 8, 2001 this sequence version replaced gi:12657099.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-409K20 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-409K20 The true
            left end of clone RP11-138E2 is at 118932 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .169963
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-409K20"
                     /clone_lib="RPCI-11.2"
     repeat_region   5. .86
                     /note="MSTC repeat: matches 46. .126 of consensus"
     misc_feature    28. .462
                     /note="match: GSS: Em:AQ718881"
     repeat_region   817. .992
                     /note="Charlie2 repeat: matches 7. .195 of consensus"
     misc_feature    complement(2510. .2941)
                     /note="match: GSS: Em:AQ041615"
     misc_feature    2944. .3096
                     /note="match: GSS: Em:B74700"
     misc_feature    3329. .4807
                     /note="CpG island"
```

```
                     /evidence=not_experimental
mRNA                 join(4205. .4464,5126. .5391,8241. .8416,9958. .10085,
                     10395. .12334)
                     /gene="TOR1B"
                     /note="match: cDNAs: Em:AF007872 Em:AJ297743
                     match: ESTs: Em:AI815528 Em:AW160403 Em:AW972065
                     Em:BF313148 Em:AA112625 Em:BE740991 Em:BE563034
                     Em:BE893335 Em:AV728123 Em:AW952051 Em:AW148938
                     Em:BE502754 Em:AW016676 Em:AI223067 Em:BE108689
                     Em:AI808893 Em:AW173267 Em:AI185247"
                     /product="bA409K20.1.1 (torsin family 1, member B (torsin
                     B) (DQ1))"
                     /evidence=not_experimental
gene                 4205. .12334
                     /gene="TOR1B"
CDS                  join(4266. .4464,5126. .5391,8241. .8416,9958. .10085,
                     10395. .10636)
                     /gene="TOR1B"
                     /note="match: proteins: Tr:O14657"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bA409K20.1.1 (torsin family 1, member B (torsin
                     B) (DQ1))"
                     /protein_id="CAC88165.1"
                     /db_xref="GI:15787707"
                     /translation="MLRAGWLRGAAALALLLAARVVAAFEPITVGLAIGAASAITGYL
                     SYNDIYCRFAECCREERPLNASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKP
                     LTLSLHGWAGTGKNFVSQIVAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQK
                     WIRGNVSACANSVFIFDEMDKLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLSNAGG
                     DLITKTALDFWRAGRKREDIQLKDLEPVLSVGVFNNKHSGLWHSGLIDKNLIDYFIPF
                     LPLEYRHVKMCVRAEMRARGSAIDEDIVTRVAEEMTFFPRDEKIYSDKGCKTVQSRLD
                     FH"
mRNA                 join(4321. .4464,5126. .5391,11280. .11571)
                     /gene="TOR1B"
                     /note="isoform 3
                     match: ESTs: Em:BF058863 Em:BE315222"
                     /product="bA409K20.1.3 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 3)"
                     /evidence=not_experimental
CDS                  join(<4321. .4464,5126. .5391,11280. .11294)
                     /gene="TOR1B"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="bA409K20.1.3 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 3)"
                     /protein_id="CAC88166.1"
                     /db_xref="GI:15787708"
                     /translation="RVVAAFEPITVGLAIGAASAITGYLSYNDIYCRFAECCREERPL
                     NASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQI
                     VAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQSSLT"
mRNA                 join(5126. .5266,8241. .8416,9958. .10085,10395. .10456)
                     /gene="TOR1B"
                     /note="match: ESTs: Em:AI468027"
                     /product="bA409K20.1.2 (isoform 2)"
                     /evidence=not_experimental
mRNA                 join(5159. .5391,8241. .8416,11280. .11319)
                     /gene="TOR1B"
```

```
                     /note="isoform 4
                     match: ESTs: Em:AI568476"
                     /product="bA409K20.1.4 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 4)"
                     /evidence=not_experimental
     CDS             join(<5159. .5391,8241. .8416,11280. .11289)
                     /gene="TOR1B"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="bA409K20.1.4 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 4)"
                     /protein_id="CAC88167.1"
                     /db_xref="GI:15787709"
                     /translation="QHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQIV
                     AENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQKWIRGNVSACANSVFIFDEMD
                     KLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLRVH"
     misc_feature    5188. .5526
                     /gene="TOR1B"
                     /note="match: STS: Em:G24606"
     repeat_region   7370. .7432
                     /note="MER61E repeat: matches 128. .190 of consensus"
     misc_feature    complement(11923. .12334)
                     /note="match: STS: Em:G27406"
     misc_feature    complement(12097. .12334)
                     /note="match: STS: Em:G24725"
     polyA_signal    12313. .12318
                     /gene="TOR1B"
     polyA_site      12334
                     /gene="TOR1B"
     polyA_site      complement(13997)
                     /gene="DYT1"
     mRNA            complement(join(13997. .15275,19573. .19700,19798. .19973,
                     23634. .23899,24961. .25180))
                     /gene="DYT1"
                     /note="match: cDNAs: Em:AF007871
                     match: ESTs: Em:BE272533 Em:BE314317 Em:BE784377
                     Em:BF203163 Em:BE622540 Em:AI039978 Em:AI770117
                     Em:AW050630 Em:AI970719 Em:BE463967 Em:AI374678
                     Em:AI167967 Em:AI127274 Em:AI699731 Em:AW001722
                     Em:AI301894 Em:AW080988"
                     /product="bA409K20.2 (dystonia 1, torsion (autosomal
                     dominant; torsin A) (DQ2, TOR1A))"
                     /evidence=not_experimental
     gene            complement(13997. .25180)
                     /gene="DYT1"
     polyA_signal    complement(14010. .14015)
                     /gene="DYT1"
     misc_feature    14016. .14298
                     /note="match: STS: Em:G30092"
     misc_feature    complement(14429. .14885)
                     /gene="DYT1"
                     /note="match: GSS: Em:B69651"
     misc_feature    complement(14469. .14876)
                     /gene="DYT1"
                     /note="match: GSS: Em:B48142"
     misc_feature    complement(14494. .14860)
                     /gene="DYT1"
```

```
                     /note="match: GSS: Em:AQ566167"
     polyA_site      complement(14632)
                     /gene="DYT1"
     misc_feature    14650. .15099
                     /note="match: STS: Em:G60041 Em:G60042"
     misc_feature    14807. .14914
                     /note="match: STS: Em:G43378 Em:G43379"
     misc_feature    14885. .15212
                     /note="match: GSS: Em:AQ213491"
     misc_feature    14890. .15392
                     /note="match: GSS: Em:AQ482600"
     CDS             complement(join(15025. .15275,19573. .19700,19798. .19973,
                     23634. .23899,24961. .25138))

  Query Match             98.0%;  Score 392;  DB 9;  Length 169963;
  Best Local Similarity   98.8%;  Pred. No. 4.3e-117;
  Matches  395;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 gaatatttacgagggtggtctgaacagtgactatgtccacctgtttgtggccacattgct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23727 GAATATTTACGAGGGTGGTCTGAACAGTGACTATGTCCACCTGTTTGTGGCCACATTGCA 23668

Qy      61 ctttccacatgcttcaaacatcaccttgtacaaggcaaggatggaagtttggaatccctt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23667 CTTTCCACATGCTTCAAACATCACCTTGTACAAGGCAAGGATGGAAGTTTGGAATCCCTT 23608

Qy     121 cctggatgtcatcgggtttggggtctctttgttgtgggatgagatttgggagttctatgt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23607 CCTGGATGTCATCGGGTTTGGGGTCTCTTTGTTGTGGGATGAGATTTGGGAGTTCTATGT 23548

Qy     181 tgaaatgagtgagcccggaaaacggttcatgtctcagttccccttggaaaggtgtagaag 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23547 TGAAATGAGTGAGCCCGGAAAACGGTTCATGTCTCAGTTCCCCTTGGAAAGGTGTAGAAG 23488

Qy     241 ttaagagtttgagatgcgtggagcagttaataccatcaaagctttgtggtgggttctgaa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23487 TTAAGAGTTTGAGATGCGTGGAGCAGTTAATACCATCAAAGCTTTGTGGTGGGTTCTGAA 23428

Qy     301 aatcggtccagtgagtatgtagggtcatgggattttagaggtggacatgatcaaatccat 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23427 AATCGGTCCAGTGAGTATGTAGGGTCATGGGATTTTAGAGGTGGACATGATCAAATCCAT 23368

Qy     361 cttagagatcaacacatctcactcatttttatttttcttat 400
          |||||||||||||||||||||||||||| | || || | |
Db   23367 CTTAGAGATCAACACATCTCACTCATTTTTTTATTTTTTT 23328
```

52

T T
I ← SID 89
C T

SID 50: 78.7% local (lots of ns) - 24142 - 24488

SID 51: 93.3% local sim 23910 - 24276

SID 53: 99.3% local sim 20257 - 19840

SID 54: 100% local sim 19859 - 19670

SID 56: 97.5 local sim 16533 - 15236

*SID 49 · 96.9% local 24658 - 24278

SID 55: 95.5% local 19601 - 19071

*SID 48 - 96.8% 24960 - 24678

*SID 88: 93.5% local sim 23900 - 24276

*SID 89: 98.5% M 23727 - 23326

SID 90: M 100% 19859 - 19660)

32-39 no 103 now.

need to click through to see when notations were available

RESULT 8
AL158207
LOCUS AL158207 169963 bp DNA linear PRI 25-SEP-2001 → This is first entry with notations.
DEFINITION Human DNA sequence from clone RP11-409K20 on chromosome 9 Contains the TOR1B gene for torsin family 1 member B (torsin B) (DQ1), the DYT1 gene for 'dystonia 1, torsion' (autosomal dominant; torsin A) (DQ2, TOR1A), the gene for hepatocellular carcinoma-associated antigen 59 (HSPC220, LOC51759), the USP20 gene for ubiquitin specific protease 20 (KIAA1003), and the gene for formin-binding protein 17 (FBP17, includes KIAA0554, FLJ13619, FLJ10754 and FLJ10113). Contains ESTs, STSs, GSSs and four CpG islands, complete sequence.
ACCESSION AL158207
VERSION AL158207.15 GI:12717949
KEYWORDS HTG; CpG island; DQ1; dystonia; DYT1; FBP17; FLJ10113; FLJ10754; FLJ13619; formin-binding; hepatocellular; HSPC220; KIAA0554; KIAA1003; LOC51759; protease; TOR1B; torsin; ubiquitin; USP20.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169963)
AUTHORS Babbage,A.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Feb 8, 2001 this sequence version replaced gi:12657099.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-409K20 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-409K20 The true left end of clone RP11-138E2 is at 118932 in this sequence.
FEATURES Location/Qualifiers

```
source          1. .169963
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="9"
                /clone="RP11-409K20"
                /clone_lib="RPCI-11.2"
repeat_region   5. .86
                /note="MSTC repeat: matches 46. .126 of consensus"
misc_feature    28. .462
                /note="match: GSS: Em:AQ718881"
repeat_region   817. .992
                /note="Charlie2 repeat: matches 7. .195 of consensus"
misc_feature    complement(2510. .2941)
                /note="match: GSS: Em:AQ041615"
misc_feature    2944. .3096
                /note="match: GSS: Em:B74700"
misc_feature    3329. .4807
                /note="CpG island"
                /evidence=not_experimental
mRNA            join(4205. .4464,5126. .5391,8241. .8416,9958. .10085,
                10395. .12334)
                /gene="TOR1B"
                /note="match: cDNAs: Em:AF007872 Em:AJ297743
                match: ESTs: Em:AI815528 Em:AW160403 Em:AW972065
                Em:BF313148 Em:AA112625 Em:BE740991 Em:BE563034
                Em:BE893335 Em:AV728123 Em:AW952051 Em:AW148938
                Em:BE502754 Em:AW016676 Em:AI223067 Em:BE108689
                Em:AI808893 Em:AW173267 Em:AI185247"
                /product="bA409K20.1.1 (torsin family 1, member B (torsin
                B) (DQ1))"
                /evidence=not_experimental
gene            4205. .12334
                /gene="TOR1B"
CDS             join(4266. .4464,5126. .5391,8241. .8416,9958. .10085,
                10395. .10636)
                /gene="TOR1B"
                /note="match: proteins: Tr:O14657"
                /codon_start=1
                /evidence=not_experimental
                /product="bA409K20.1.1 (torsin family 1, member B (torsin
                B) (DQ1))"
                /protein_id="CAC88165.1"
                /db_xref="GI:15787707"
                /translation="MLRAGWLRGAAALALLLAARVVAAFEPITVGLAIGAASAITGYL
                SYNDIYCRFAECCREERPLNASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKP
                LTLSLHGWAGTGKNFVSQIVAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQK
                WIRGNVSACANSVFIFDEMDKLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLSNAGG
                DLITKTALDFWRAGRKREDIQLKDLEPVLSVGVFNNKHSGLWHSGLIDKNLIDYFIPF
                LPLEYRHVKMCVRAEMRARGSAIDEDIVTRVAEEMTFFPRDEKIYSDKGCKTVQSRLD
                FH"
mRNA            join(4321. .4464,5126. .5391,11280. .11571)
                /gene="TOR1B"
                /note="isoform 3
                match: ESTs: Em:BF058863 Em:BE315222"
                /product="bA409K20.1.3 (torsin family 1, member B (torsin
                B) (DQ1), putative isoform 3)"
                /evidence=not_experimental
```

```
CDS             join(<4321. .4464,5126. .5391,11280. .11294)
                /gene="TOR1B"
                /codon_start=3
                /evidence=not_experimental
                /product="bA409K20.1.3 (torsin family 1, member B (torsin
                B) (DQ1), putative isoform 3)"
                /protein_id="CAC88166.1"
                /db_xref="GI:15787708"
                /translation="RVVAAFEPITVGLAIGAASAITGYLSYNDIYCRFAECCREERPL
                NASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQI
                VAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQSSLT"
mRNA            join(5126. .5266,8241. .8416,9958. .10085,10395. .10456)
                /gene="TOR1B"
                /note="match: ESTs: Em:AI468027"
                /product="bA409K20.1.2 (isoform 2)"
                /evidence=not_experimental
mRNA            join(5159. .5391,8241. .8416,11280. .11319)
                /gene="TOR1B"
                /note="isoform 4
                match: ESTs: Em:AI568476"
                /product="bA409K20.1.4 (torsin family 1, member B (torsin
                B) (DQ1), putative isoform 4)"
                /evidence=not_experimental
CDS             join(<5159. .5391,8241. .8416,11280. .11289)
                /gene="TOR1B"
                /codon_start=3
                /evidence=not_experimental
                /product="bA409K20.1.4 (torsin family 1, member B (torsin
                B) (DQ1), putative isoform 4)"
                /protein_id="CAC88167.1"
                /db_xref="GI:15787709"
                /translation="QHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQIV
                AENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQKWIRGNVSACANSVFIFDEMD
                KLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLRVH"
misc_feature    5188. .5526
                /gene="TOR1B"
                /note="match: STS: Em:G24606"
repeat_region   7370. .7432
                /note="MER61E repeat: matches 128. .190 of consensus"
misc_feature    complement(11923. .12334)
                /note="match: STS: Em:G27406"
misc_feature    complement(12097. .12334)
                /note="match: STS: Em:G24725"
polyA_signal    12313. .12318
                /gene="TOR1B"
polyA_site      12334
                /gene="TOR1B"
polyA_site      complement(13997)
                /gene="DYT1"
mRNA            complement(join(13997. .15275,19573. .19700,19798. .19973,
                23634. .23899,24961. .25180))
                /gene="DYT1"
                /note="match: cDNAs: Em:AF007871
                match: ESTs: Em:BE272533 Em:BE314317 Em:BE784377
                Em:BF203163 Em:BE622540 Em:AI039978 Em:AI770117
                Em:AW050630 Em:AI970719 Em:BE463967 Em:AI374678
                Em:AI167967 Em:AI127274 Em:AI699731 Em:AW001722
```

```
                     Em:AI301894 Em:AW080988"
                     /product="bA409K20.2 (dystonia 1, torsion (autosomal
                     dominant; torsin A) (DQ2, TOR1A))"
                     /evidence=not_experimental
     gene            complement(13997. .25180)
                     /gene="DYT1"
     polyA_signal    complement(14010. .14015)
                     /gene="DYT1"
     misc_feature    14016. .14298
                     /note="match: STS: Em:G30092"
     misc_feature    complement(14429. .14885)
                     /gene="DYT1"
                     /note="match: GSS: Em:B69651"
     misc_feature    complement(14469. .14876)
                     /gene="DYT1"
                     /note="match: GSS: Em:B48142"
     misc_feature    complement(14494. .14860)
                     /gene="DYT1"
                     /note="match: GSS: Em:AQ566167"
     polyA_site      complement(14632)
                     /gene="DYT1"
     misc_feature    14650. .15099
                     /note="match: STS: Em:G60041 Em:G60042"
     misc_feature    14807. .14914
                     /note="match: STS: Em:G43378 Em:G43379"
     misc_feature    14885. .15212
                     /note="match: GSS: Em:AQ213491"
     misc_feature    14890. .15392
                     /note="match: GSS: Em:AQ482600"
     CDS             complement(join(15025. .15275,19573. .19700,19798. .19973,
                     23634. .23899,24961. .25138))

  Query Match             100.0%;  Score 20;  DB 9;  Length 169963;
  Best Local Similarity   100.0%;  Pred. No. 9.2;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 agtagagacgcgggtagatg 20
           ||||||||||||||||||||
Db   25009 AGTAGAGACGCGGGTAGATG 25028
```

SID 31

25019 - 25000 - SID 32

24019 - 24000 - SID 34

24880 - 24899 SID 33

23
25305 - 25285 SID 30

23610 - 23629 - SID 35
20135 - 20116 - SID 36
19332 - 19353 - SID 37
15390 - 15371 - SID 38
14751 - 14771 - SID 39

SID 54

RESULT 1
AC027008
LOCUS AC027008 166889 bp DNA linear HTG 08-NOV-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-212N14 map 8, WORKING DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION AC027008
VERSION AC027008.4 GI:10280898
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166889)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-212N14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166889)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 23, 2000 this sequence version replaced gi:8080819.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L8771
Center clone name: 212_N_14
---------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 150653 bases at least Q40
Consensus quality: 158747 bases at least Q30
Consensus quality: 161880 bases at least Q20
Insert size: 186000; agarose-fp
Insert size: 164389; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 140: contig of 140 bp in length
* 141 240: gap of 100 bp
* 241 1566: contig of 1326 bp in length
* 1567 1666: gap of 100 bp
* 1667 26279: contig of 24613 bp in length
* 26280 26379: gap of 100 bp
* 26380 27676: contig of 1297 bp in length
* 27677 27776: gap of 100 bp
* 27777 29820: contig of 2044 bp in length
* 29821 29920: gap of 100 bp
* 29921 33216: contig of 3296 bp in length
* 33217 33316: gap of 100 bp
* 33317 36627: contig of 3311 bp in length
* 36628 36727: gap of 100 bp
* 36728 39382: contig of 2655 bp in length
* 39383 39482: gap of 100 bp
* 39483 42417: contig of 2935 bp in length
* 42418 42517: gap of 100 bp
* 42518 46306: contig of 3789 bp in length
* 46307 46406: gap of 100 bp
* 46407 50207: contig of 3801 bp in length
* 50208 50307: gap of 100 bp
* 50308 53363: contig of 3056 bp in length
* 53364 53463: gap of 100 bp
* 53464 56760: contig of 3297 bp in length
* 56761 56860: gap of 100 bp
* 56861 61207: contig of 4347 bp in length
* 61208 61307: gap of 100 bp
* 61308 65984: contig of 4677 bp in length
* 65985 66084: gap of 100 bp
* 66085 72072: contig of 5988 bp in length
* 72073 72172: gap of 100 bp
* 72173 77741: contig of 5569 bp in length
* 77742 77841: gap of 100 bp
* 77842 85850: contig of 8009 bp in length
* 85851 85950: gap of 100 bp
* 85951 92902: contig of 6952 bp in length
* 92903 93002: gap of 100 bp
* 93003 103668: contig of 10666 bp in length
* 103669 103768: gap of 100 bp
* 103769 109322: contig of 5554 bp in length

```
            *    109323 109422: gap of      100 bp
            *    109423   118526: contig of 9104 bp in length
            *    118527 118626: gap of      100 bp
            *    118627   128874: contig of 10248 bp in length
            *    128875 128974: gap of      100 bp
            *    128975   138016: contig of 9042 bp in length
            *    138017 138116: gap of      100 bp
            *    138117   166500: contig of 28384 bp in length
            *    166501 166600: gap of      100 bp
            *    166601   166889: contig of 289 bp in length.
FEATURES             Location/Qualifiers
     source          1. .166889
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-212N14"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .140
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    241. .1566
                     /note="assembly_fragment"
     misc_feature    1667. .26279
                     /note="assembly_fragment"
     misc_feature    26380. .27676
                     /note="assembly_fragment"
     misc_feature    27777. .29820
                     /note="assembly_fragment"
     misc_feature    29921. .33216
                     /note="assembly_fragment"
     misc_feature    33317. .36627
                     /note="assembly_fragment"
     misc_feature    36728. .39382
                     /note="assembly_fragment"
     misc_feature    39483. .42417
                     /note="assembly_fragment"
     misc_feature    42518. .46306
                     /note="assembly_fragment"
     misc_feature    46407. .50207
                     /note="assembly_fragment"
     misc_feature    50308. .53363
                     /note="assembly_fragment"
     misc_feature    53464. .56760
                     /note="assembly_fragment"
     misc_feature    56861. .61207
                     /note="assembly_fragment"
     misc_feature    61308. .65984
                     /note="assembly_fragment"
     misc_feature    66085. .72072
                     /note="assembly_fragment"
     misc_feature    72173. .77741
                     /note="assembly_fragment"
     misc_feature    77842. .85850
                     /note="assembly_fragment"
     misc_feature    85951. .92902
```

```
                     /note="assembly_fragment"
     misc_feature    93003. .103668
                     /note="assembly_fragment"
     misc_feature    103769. .109322
                     /note="assembly_fragment"
     misc_feature    109423. .118526
                     /note="assembly_fragment"
     misc_feature    118627. .128874
                     /note="assembly_fragment"
     misc_feature    128975. .138016
                     /note="assembly_fragment"
     misc_feature    138117. .166500
                     /note="assembly_fragment"
     misc_feature    166601. .166889
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    43782 a  38337 c  39037 g  43225 t   2508 others
ORIGIN


  Query Match             96.0%;  Score 190;  DB 2;  Length 166889;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-49;
  Matches  190;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctcgactattatgacctggtggatggggtctcctaccagaaagccatgttcatatttctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104685 CTCGACTATTATGACCTGGTGGATGGGGTCTCCTACCAGAAAGCCATGTTCATATTTCTC 104744

Qy     61 aggtaaggtcagggctaggacatgatggatgggccccgagcccaagcctctgagctccag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104745 AGGTAAGGTCAGGGCTAGGACATGATGGATGGGCCCCGAGCCCAAGCCTCTGAGCTCCAG 104804

Qy    121 gagaaaaccctgtccttacccactgggattgttttgcagcaatgctggagcagaaaggat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104805 GAGAAAACCCTGTCCTTACCCACTGGGATTGTTTTGCAGCAATGCTGGAGCAGAAAGGAT 104864

Qy    181 cacagatgtg 190
          ||||||||||
Db 104865 CACAGATGTG 104874


RESULT   2
AL158207/c
LOCUS       AL158207               169963 bp    DNA     linear   PRI 25-SEP-2001
DEFINITION  Human DNA sequence from clone RP11-409K20 on chromosome 9 Contains
            the TOR1B gene for torsin family 1 member B (torsin B) (DQ1), the
            DYT1 gene for 'dystonia 1, torsion' (autosomal dominant; torsin A)
            (DQ2, TOR1A), the gene for hepatocellular carcinoma-associated
            antigen 59 (HSPC220, LOC51759), the USP20 gene for ubiquitin
            specific protease 20 (KIAA1003), and the gene for formin-binding
            protein 17 (FBP17, includes KIAA0554, FLJ13619, FLJ10754 and
            FLJ10113). Contains ESTs, STSs, GSSs and four CpG islands, complete
            sequence.
ACCESSION   AL158207
VERSION     AL158207.15  GI:12717949
KEYWORDS    HTG; CpG island; DQ1; dystonia; DYT1; FBP17; FLJ10113; FLJ10754;
```

FLJ13619; formin-binding; hepatocellular; HSPC220; KIAA0554; KIAA1003; LOC51759; protease; TOR1B; torsin; ubiquitin; USP20.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 169963)
AUTHORS Babbage,A.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT On Feb 8, 2001 this sequence version replaced gi:12657099.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-409K20 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-409K20 The true left end of clone RP11-138E2 is at 118932 in this sequence.

FEATURES Location/Qualifiers

source	1. .169963
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="9"
	/clone="RP11-409K20"
	/clone_lib="RPCI-11.2"
repeat_region	5. .86
	/note="MSTC repeat: matches 46. .126 of consensus"
misc_feature	28. .462
	/note="match: GSS: Em:AQ718881"
repeat_region	817. .992
	/note="Charlie2 repeat: matches 7. .195 of consensus"
misc_feature	complement(2510. .2941)
	/note="match: GSS: Em:AQ041615"
misc_feature	2944. .3096

```
                     /note="match: GSS: Em:B74700"
misc_feature         3329. .4807
                     /note="CpG island"
                     /evidence=not_experimental
mRNA                 join(4205. .4464,5126. .5391,8241. .8416,9958. .10085,
                     10395. .12334)
                     /gene="TOR1B"
                     /note="match: cDNAs: Em:AF007872 Em:AJ297743
                     match: ESTs: Em:AI815528 Em:AW160403 Em:AW972065
                     Em:BF313148 Em:AA112625 Em:BE740991 Em:BE563034
                     Em:BE893335 Em:AV728123 Em:AW952051 Em:AW148938
                     Em:BE502754 Em:AW016676 Em:AI223067 Em:BE108689
                     Em:AI808893 Em:AW173267 Em:AI185247"
                     /product="bA409K20.1.1 (torsin family 1, member B (torsin
                     B) (DQ1))"
                     /evidence=not_experimental
gene                 4205. .12334
                     /gene="TOR1B"
CDS                  join(4266. .4464,5126. .5391,8241. .8416,9958. .10085,
                     10395. .10636)
                     /gene="TOR1B"
                     /note="match: proteins: Tr:O14657"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bA409K20.1.1 (torsin family 1, member B (torsin
                     B) (DQ1))"
                     /protein_id="CAC88165.1"
                     /db_xref="GI:15787707"
                     /translation="MLRAGWLRGAAALALLLAARVVAAFEPITVGLAIGAASAITGYL
                     SYNDIYCRFAECCREERPLNASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKP
                     LTLSLHGWAGTGKNFVSQIVAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQK
                     WIRGNVSACANSVFIFDEMDKLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLSNAGG
                     DLITKTALDFWRAGRKREDIQLKDLEPVLSVGVFNNKHSGLWHSGLIDKNLIDYFIPF
                     LPLEYRHVKMCVRAEMRARGSAIDEDIVTRVAEEMTFFPRDEKIYSDKGCKTVQSRLD
                     FH"
mRNA                 join(4321. .4464,5126. .5391,11280. .11571)
                     /gene="TOR1B"
                     /note="isoform 3
                     match: ESTs: Em:BF058863 Em:BE315222"
                     /product="bA409K20.1.3 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 3)"
                     /evidence=not_experimental
CDS                  join(<4321. .4464,5126. .5391,11280. .11294)
                     /gene="TOR1B"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="bA409K20.1.3 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 3)"
                     /protein_id="CAC88166.1"
                     /db_xref="GI:15787708"
                     /translation="RVVAAFEPITVGLAIGAASAITGYLSYNDIYCRFAECCREERPL
                     NASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQI
                     VAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQSSLT"
mRNA                 join(5126. .5266,8241. .8416,9958. .10085,10395. .10456)
                     /gene="TOR1B"
                     /note="match: ESTs: Em:AI468027"
                     /product="bA409K20.1.2 (isoform 2)"
```

```
                     /evidence=not_experimental
mRNA                 join(5159. .5391,8241. .8416,11280. .11319)
                     /gene="TOR1B"
                     /note="isoform 4
                     match: ESTs: Em:AI568476"
                     /product="bA409K20.1.4 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 4)"
                     /evidence=not_experimental
CDS                  join(<5159. .5391,8241. .8416,11280. .11289)
                     /gene="TOR1B"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="bA409K20.1.4 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 4)"
                     /protein_id="CAC88167.1"
                     /db_xref="GI:15787709"
                     /translation="QHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQIV
                     AENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQKWIRGNVSACANSVFIFDEMD
                     KLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLRVH"
misc_feature         5188. .5526
                     /gene="TOR1B"
                     /note="match: STS: Em:G24606"
repeat_region        7370. .7432
                     /note="MER61E repeat: matches 128. .190 of consensus"
misc_feature         complement(11923. .12334)
                     /note="match: STS: Em:G27406"
misc_feature         complement(12097. .12334)
                     /note="match: STS: Em:G24725"
polyA_signal         12313. .12318
                     /gene="TOR1B"
polyA_site           12334
                     /gene="TOR1B"
polyA_site           complement(13997)
                     /gene="DYT1"
mRNA                 complement(join(13997. .15275,19573. .19700,19798. .19973,
                     23634. .23899,24961. .25180))
                     /gene="DYT1"
                     /note="match: cDNAs: Em:AF007871
                     match: ESTs: Em:BE272533 Em:BE314317 Em:BE784377
                     Em:BF203163 Em:BE622540 Em:AI039978 Em:AI770117
                     Em:AW050630 Em:AI970719 Em:BE463967 Em:AI374678
                     Em:AI167967 Em:AI127274 Em:AI699731 Em:AW001722
                     Em:AI301894 Em:AW080988"
                     /product="bA409K20.2 (dystonia 1, torsion (autosomal
                     dominant; torsin A) (DQ2, TOR1A))"
                     /evidence=not_experimental
gene                 complement(13997. .25180)
                     /gene="DYT1"
polyA_signal         complement(14010. .14015)
                     /gene="DYT1"
misc_feature         14016. .14298
                     /note="match: STS: Em:G30092"
misc_feature         complement(14429. .14885)
                     /gene="DYT1"
                     /note="match: GSS: Em:B69651"
misc_feature         complement(14469. .14876)
                     /gene="DYT1"
```

```
                     /note="match: GSS: Em:B48142"
     misc_feature    complement(14494. .14860)
                     /gene="DYT1"
                     /note="match: GSS: Em:AQ566167"
     polyA_site      complement(14632)
                     /gene="DYT1"
     misc_feature    14650. .15099
                     /note="match: STS: Em:G60041 Em:G60042"
     misc_feature    14807. .14914
                     /note="match: STS: Em:G43378 Em:G43379"
     misc_feature    14885. .15212
                     /note="match: GSS: Em:AQ213491"
     misc_feature    14890. .15392
                     /note="match: GSS: Em:AQ482600"
     CDS             complement(join(15025. .15275,19573. .19700,19798. .19973,
                     23634. .23899,24961. .25138))

  Query Match             96.0%;  Score 190;  DB 9;  Length 169963;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-49;
  Matches  190;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ctcgactattatgacctggtggatggggtctcctaccagaaagccatgttcatatttctc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19859 CTCGACTATTATGACCTGGTGGATGGGGTCTCCTACCAGAAAGCCATGTTCATATTTCTC 19800

Qy     61 aggtaaggtcagggctaggacatgatggatgggccccgagcccaagcctctgagctccag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19799 AGGTAAGGTCAGGGCTAGGACATGATGGATGGGCCCCGAGCCCAAGCCTCTGAGCTCCAG 19740

Qy    121 gagaaaaccctgtccttacccactgggattgttttgcagcaatgctggagcagaaaggat 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19739 GAGAAAACCCTGTCCTTACCCACTGGGATTGTTTTGCAGCAATGCTGGAGCAGAAAGGAT 19680

Qy    181 cacagatgtg 190
          ||||||||||
Db  19679 CACAGATGTG 19670
```

Prior art SID 53 against identity clms

RESULT 1
AC027008
LOCUS AC027008 166889 bp DNA linear HTG 08-NOV-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-212N14 map 8, WORKING DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION AC027008
VERSION AC027008.4 GI:10280898
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166889)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-212N14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166889)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 23, 2000 this sequence version replaced gi:8080819.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L8771
Center clone name: 212_N_14
---------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 150653 bases at least Q40
Consensus quality: 158747 bases at least Q30
Consensus quality: 161880 bases at least Q20
Insert size: 186000; agarose-fp
Insert size: 164389; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 140: contig of 140 bp in length
* 141 240: gap of 100 bp
* 241 1566: contig of 1326 bp in length
* 1567 1666: gap of 100 bp
* 1667 26279: contig of 24613 bp in length
* 26280 26379: gap of 100 bp
* 26380 27676: contig of 1297 bp in length
* 27677 27776: gap of 100 bp
* 27777 29820: contig of 2044 bp in length
* 29821 29920: gap of 100 bp
* 29921 33216: contig of 3296 bp in length
* 33217 33316: gap of 100 bp
* 33317 36627: contig of 3311 bp in length
* 36628 36727: gap of 100 bp
* 36728 39382: contig of 2655 bp in length
* 39383 39482: gap of 100 bp
* 39483 42417: contig of 2935 bp in length
* 42418 42517: gap of 100 bp
* 42518 46306: contig of 3789 bp in length
* 46307 46406: gap of 100 bp
* 46407 50207: contig of 3801 bp in length
* 50208 50307: gap of 100 bp
* 50308 53363: contig of 3056 bp in length
* 53364 53463: gap of 100 bp
* 53464 56760: contig of 3297 bp in length
* 56761 56860: gap of 100 bp
* 56861 61207: contig of 4347 bp in length
* 61208 61307: gap of 100 bp
* 61308 65984: contig of 4677 bp in length
* 65985 66084: gap of 100 bp
* 66085 72072: contig of 5988 bp in length
* 72073 72172: gap of 100 bp
* 72173 77741: contig of 5569 bp in length
* 77742 77841: gap of 100 bp
* 77842 85850: contig of 8009 bp in length
* 85851 85950: gap of 100 bp
* 85951 92902: contig of 6952 bp in length
* 92903 93002: gap of 100 bp
* 93003 103668: contig of 10666 bp in length
* 103669 103768: gap of 100 bp
* 103769 109322: contig of 5554 bp in length

```
            *    109323 109422: gap of      100 bp
            *    109423   118526: contig of 9104 bp in length
            *    118527 118626: gap of      100 bp
            *    118627   128874: contig of 10248 bp in length
            *    128875 128974: gap of      100 bp
            *    128975   138016: contig of 9042 bp in length
            *    138017 138116: gap of      100 bp
            *    138117   166500: contig of 28384 bp in length
            *    166501 166600: gap of      100 bp
            *    166601   166889: contig of 289 bp in length.
FEATURES             Location/Qualifiers
     source          1. .166889
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-212N14"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .140
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    241. .1566
                     /note="assembly_fragment"
     misc_feature    1667. .26279
                     /note="assembly_fragment"
     misc_feature    26380. .27676
                     /note="assembly_fragment"
     misc_feature    27777. .29820
                     /note="assembly_fragment"
     misc_feature    29921. .33216
                     /note="assembly_fragment"
     misc_feature    33317. .36627
                     /note="assembly_fragment"
     misc_feature    36728. .39382
                     /note="assembly_fragment"
     misc_feature    39483. .42417
                     /note="assembly_fragment"
     misc_feature    42518. .46306
                     /note="assembly_fragment"
     misc_feature    46407. .50207
                     /note="assembly_fragment"
     misc_feature    50308. .53363
                     /note="assembly_fragment"
     misc_feature    53464. .56760
                     /note="assembly_fragment"
     misc_feature    56861. .61207
                     /note="assembly_fragment"
     misc_feature    61308. .65984
                     /note="assembly_fragment"
     misc_feature    66085. .72072
                     /note="assembly_fragment"
     misc_feature    72173. .77741
                     /note="assembly_fragment"
     misc_feature    77842. .85850
                     /note="assembly_fragment"
     misc_feature    85951. .92902
```

```
                     /note="assembly_fragment"
     misc_feature    93003. .103668
                     /note="assembly_fragment"
     misc_feature    103769. .109322
                     /note="assembly_fragment"
     misc_feature    109423. .118526
                     /note="assembly_fragment"
     misc_feature    118627. .128874
                     /note="assembly_fragment"
     misc_feature    128975. .138016
                     /note="assembly_fragment"
     misc_feature    138117. .166500
                     /note="assembly_fragment"
     misc_feature    166601. .166889
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    43782 a  38337 c  39037 g  43225 t   2508 others
ORIGIN


  Query Match             99.3%;  Score 415;  DB 2;  Length 166889;
  Best Local Similarity   99.3%;  Pred. No. 4.1e-111;
  Matches  415;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 tttggagtgagacaggactgggttcaggtcccagctctgccacatatagtcttgggcaag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104287 TTTGGAGTGAGACAGGACTGGGTTCAGGTCCCAGCTCTGCCACATATAGTCTTGGGCAAG 104346

Qy     61 tggagtaagcgctctctgtgcctcagttccctcatctgtaaaatgagaacgatagtgccc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104347 TGGAGTAAGCGCTCTCTGTGCCTCAGTTCCCTCATCTGTAAAATGAGAACGATAGTGCCC 104406

Qy    121 actccatggggttggtaggaacaaagaagattttgggcatgtaaagttcttagtgccgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104407 ACTCCATGGGGTTGGTAGGAACAAAGAAGATTTTGGGCATGTAAAGTTCTTAGTGCCGAG 104466

Qy    181 tgcacagtggtctgtaagtgaagctgcggttcttagtggtagaaggagctgattgatggc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104467 TGCACAGTGGTCTGTAAGTGAAGCTGCGGTTCTTAGTGGTAGAAGGAGCTGATTGATGGC 104526

Qy    241 cctggctgagaactttgtgttcgctttttcccntttttaattcaggatcagttacagttgt 300
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 104527 CCTGGCTGAGAACTTTGTGTTCGCTTTTTCCCCTTTTAATTCAGGATCAGTTACAGTTGT 104586

Qy    301 ggattcgaggcaacgtgagtgcctgtgcgaggtccatcttcatatttgatgaaatggata 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 104587 GGATTCGAGGCAACGTGAGTGCCTGTGCGAGGTCCATCTTCATATTTGATGAAATGGATA 104646

Qy    361 agatgcatgcaggcctcatagatgccntcaancctttcctcgactattatgacctggt 418
          |||||||||||||||||||||||||| |||| ||||||||||||||||||||||||||
Db 104647 AGATGCATGCAGGCCTCATAGATGCCATCAAGCCTTTCCTCGACTATTATGACCTGGT 104704


RESULT   2
AL158207/c
LOCUS       AL158207              169963 bp    DNA     linear   PRI 25-SEP-2001
```

DEFINITION Human DNA sequence from clone RP11-409K20 on chromosome 9 Contains the TOR1B gene for torsin family 1 member B (torsin B) (DQ1), the DYT1 gene for 'dystonia 1, torsion' (autosomal dominant; torsin A) (DQ2, TOR1A), the gene for hepatocellular carcinoma-associated antigen 59 (HSPC220, LOC51759), the USP20 gene for ubiquitin specific protease 20 (KIAA1003), and the gene for formin-binding protein 17 (FBP17, includes KIAA0554, FLJ13619, FLJ10754 and FLJ10113). Contains ESTs, STSs, GSSs and four CpG islands, complete sequence.
ACCESSION AL158207
VERSION AL158207.15 GI:12717949
KEYWORDS HTG; CpG island; DQ1; dystonia; DYT1; FBP17; FLJ10113; FLJ10754; FLJ13619; formin-binding; hepatocellular; HSPC220; KIAA0554; KIAA1003; LOC51759; protease; TOR1B; torsin; ubiquitin; USP20.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169963)
AUTHORS Babbage,A.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Feb 8, 2001 this sequence version replaced gi:12657099.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-409K20 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-409K20 The true left end of clone RP11-138E2 is at 118932 in this sequence.

FEATURES Location/Qualifiers
source 1. .169963
/organism="Homo sapiens"
/db_xref="taxon:9606"

```
                     /chromosome="9"
                     /clone="RP11-409K20"
                     /clone_lib="RPCI-11.2"
repeat_region        5. .86
                     /note="MSTC repeat: matches 46. .126 of consensus"
misc_feature         28. .462
                     /note="match: GSS: Em:AQ718881"
repeat_region        817. .992
                     /note="Charlie2 repeat: matches 7. .195 of consensus"
misc_feature         complement(2510. .2941)
                     /note="match: GSS: Em:AQ041615"
misc_feature         2944. .3096
                     /note="match: GSS: Em:B74700"
misc_feature         3329. .4807
                     /note="CpG island"
                     /evidence=not_experimental
mRNA                 join(4205. .4464,5126. .5391,8241. .8416,9958. .10085,
                     10395. .12334)
                     /gene="TOR1B"
                     /note="match: cDNAs: Em:AF007872 Em:AJ297743
                     match: ESTs: Em:AI815528 Em:AW160403 Em:AW972065
                     Em:BF313148 Em:AA112625 Em:BE740991 Em:BE563034
                     Em:BE893335 Em:AV728123 Em:AW952051 Em:AW148938
                     Em:BE502754 Em:AW016676 Em:AI223067 Em:BE108689
                     Em:AI808893 Em:AW173267 Em:AI185247"
                     /product="bA409K20.1.1 (torsin family 1, member B (torsin
                     B) (DQ1))"
                     /evidence=not_experimental
gene                 4205. .12334
                     /gene="TOR1B"
CDS                  join(4266. .4464,5126. .5391,8241. .8416,9958. .10085,
                     10395. .10636)
                     /gene="TOR1B"
                     /note="match: proteins: Tr:O14657"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bA409K20.1.1 (torsin family 1, member B (torsin
                     B) (DQ1))"
                     /protein_id="CAC88165.1"
                     /db_xref="GI:15787707"
                     /translation="MLRAGWLRGAAALALLLAARVVAAFEPITVGLAIGAASAITGYL
                     SYNDIYCRFAECCREERPLNASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKP
                     LTLSLHGWAGTGKNFVSQIVAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQK
                     WIRGNVSACANSVFIFDEMDKLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLSNAGG
                     DLITKTALDFWRAGRKREDIQLKDLEPVLSVGVFNNKHSGLWHSGLIDKNLIDYFIPF
                     LPLEYRHVKMCVRAEMRARGSAIDEDIVTRVAEEMTFFPRDEKIYSDKGCKTVQSRLD
                     FH"
mRNA                 join(4321. .4464,5126. .5391,11280. .11571)
                     /gene="TOR1B"
                     /note="isoform 3
                     match: ESTs: Em:BF058863 Em:BE315222"
                     /product="bA409K20.1.3 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 3)"
                     /evidence=not_experimental
CDS                  join(<4321. .4464,5126. .5391,11280. .11294)
                     /gene="TOR1B"
                     /codon_start=3
```

```
                     /evidence=not_experimental
                     /product="bA409K20.1.3 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 3)"
                     /protein_id="CAC88166.1"
                     /db_xref="GI:15787708"
                     /translation="RVVAAFEPITVGLAIGAASAITGYLSYNDIYCRFAECCREERPL
                     NASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQI
                     VAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQSSLT"
mRNA                 join(5126. .5266,8241. .8416,9958. .10085,10395. .10456)
                     /gene="TOR1B"
                     /note="match: ESTs: Em:AI468027"
                     /product="bA409K20.1.2 (isoform 2)"
                     /evidence=not_experimental
mRNA                 join(5159. .5391,8241. .8416,11280. .11319)
                     /gene="TOR1B"
                     /note="isoform 4
                     match: ESTs: Em:AI568476"
                     /product="bA409K20.1.4 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 4)"
                     /evidence=not_experimental
CDS                  join(<5159. .5391,8241. .8416,11280. .11289)
                     /gene="TOR1B"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="bA409K20.1.4 (torsin family 1, member B (torsin
                     B) (DQ1), putative isoform 4)"
                     /protein_id="CAC88167.1"
                     /db_xref="GI:15787709"
                     /translation="QHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQIV
                     AENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQKWIRGNVSACANSVFIFDEMD
                     KLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLRVH"
misc_feature         5188. .5526
                     /gene="TOR1B"
                     /note="match: STS: Em:G24606"
repeat_region        7370. .7432
                     /note="MER61E repeat: matches 128. .190 of consensus"
misc_feature         complement(11923. .12334)
                     /note="match: STS: Em:G27406"
misc_feature         complement(12097. .12334)
                     /note="match: STS: Em:G24725"
polyA_signal         12313. .12318
                     /gene="TOR1B"
polyA_site           12334
                     /gene="TOR1B"
polyA_site           complement(13997)
                     /gene="DYT1"
mRNA                 complement(join(13997. .15275,19573. .19700,19798. .19973,
                     23634. .23899,24961. .25180))
                     /gene="DYT1"
                     /note="match: cDNAs: Em:AF007871
                     match: ESTs: Em:BE272533 Em:BE314317 Em:BE784377
                     Em:BF203163 Em:BE622540 Em:AI039978 Em:AI770117
                     Em:AW050630 Em:AI970719 Em:BE463967 Em:AI374678
                     Em:AI167967 Em:AI127274 Em:AI699731 Em:AW001722
                     Em:AI301894 Em:AW080988"
                     /product="bA409K20.2 (dystonia 1, torsion (autosomal
                     dominant; torsin A) (DQ2, TOR1A))"
```

```
                     /evidence=not_experimental
     gene            complement(13997. .25180)
                     /gene="DYT1"
     polyA_signal    complement(14010. .14015)
                     /gene="DYT1"
     misc_feature    14016. .14298
                     /note="match: STS: Em:G30092"
     misc_feature    complement(14429. .14885)
                     /gene="DYT1"
                     /note="match: GSS: Em:B69651"
     misc_feature    complement(14469. .14876)
                     /gene="DYT1"
                     /note="match: GSS: Em:B48142"
     misc_feature    complement(14494. .14860)
                     /gene="DYT1"
                     /note="match: GSS: Em:AQ566167"
     polyA_site      complement(14632)
                     /gene="DYT1"
     misc_feature    14650. .15099
                     /note="match: STS: Em:G60041 Em:G60042"
     misc_feature    14807. .14914
                     /note="match: STS: Em:G43378 Em:G43379"
     misc_feature    14885. .15212
                     /note="match: GSS: Em:AQ213491"
     misc_feature    14890. .15392
                     /note="match: GSS: Em:AQ482600"
     CDS             complement(join(15025. .15275,19573. .19700,19798. .19973,
                     23634. .23899,24961. .25138))

  Query Match             99.3%;  Score 415;  DB 9;  Length 169963;
  Best Local Similarity   99.3%;  Pred. No. 4.1e-111;
  Matches  415;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 tttggagtgagacaggactgggttcaggtcccagctctgccacatatagtcttgggcaag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20257 TTTGGAGTGAGACAGGACTGGGTTCAGGTCCCAGCTCTGCCACATATAGTCTTGGGCAAG 20198

Qy     61 tggagtaagcgctctctgtgcctcagttccctcatctgtaaaatgagaacgatagtgccc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20197 TGGAGTAAGCGCTCTCTGTGCCTCAGTTCCCTCATCTGTAAAATGAGAACGATAGTGCCC 20138

Qy    121 actccatggggttggtaggaacaaagaagattttgggcatgtaaagttcttagtgccgag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20137 ACTCCATGGGGTTGGTAGGAACAAAGAAGATTTTGGGCATGTAAAGTTCTTAGTGCCGAG 20078

Qy    181 tgcacagtggtctgtaagtgaagctgcggttcttagtggtagaaggagctgattgatggc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20077 TGCACAGTGGTCTGTAAGTGAAGCTGCGGTTCTTAGTGGTAGAAGGAGCTGATTGATGGC 20018

Qy    241 cctggctgagaactttgtgttcgctttttcccntttttaattcaggatcagttacagttgt 300
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  20017 CCTGGCTGAGAACTTTGTGTTCGCTTTTTCCCCTTTTAATTCAGGATCAGTTACAGTTGT 19958

Qy    301 ggattcgaggcaacgtgagtgcctgtgcgaggtccatcttcatatttgatgaaatggata 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19957 GGATTCGAGGCAACGTGAGTGCCTGTGCGAGGTCCATCTTCATATTTGATGAAATGGATA 19898
```

```
Qy    361 agatgcatgcaggcctcatagatgccntcaanccttcctcgactattatgacctggt 418
          ||||||||||||||||||||||||| |||| ||||||||||||||||||||||||||
Db  19897 AGATGCATGCAGGCCTCATAGATGCCATCAAGCCTTTCCTCGACTATTATGACCTGGT 19840
```

SID 56 against homology DB

RESULT 1
AC027008
LOCUS AC027008 166889 bp DNA linear HTG 08-NOV-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-212N14 map 8, WORKING DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION AC027008
VERSION AC027008.4 GI:10280898
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166889)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-212N14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166889)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 23, 2000 this sequence version replaced gi:8080819.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L8771
Center clone name: 212_N_14
---------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 150653 bases at least Q40
Consensus quality: 158747 bases at least Q30
Consensus quality: 161880 bases at least Q20
Insert size: 186000; agarose-fp
Insert size: 164389; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 140: contig of 140 bp in length
* 141 240: gap of 100 bp
* 241 1566: contig of 1326 bp in length
* 1567 1666: gap of 100 bp
* 1667 26279: contig of 24613 bp in length
* 26280 26379: gap of 100 bp
* 26380 27676: contig of 1297 bp in length
* 27677 27776: gap of 100 bp
* 27777 29820: contig of 2044 bp in length
* 29821 29920: gap of 100 bp
* 29921 33216: contig of 3296 bp in length
* 33217 33316: gap of 100 bp
* 33317 36627: contig of 3311 bp in length
* 36628 36727: gap of 100 bp
* 36728 39382: contig of 2655 bp in length
* 39383 39482: gap of 100 bp
* 39483 42417: contig of 2935 bp in length
* 42418 42517: gap of 100 bp
* 42518 46306: contig of 3789 bp in length
* 46307 46406: gap of 100 bp
* 46407 50207: contig of 3801 bp in length
* 50208 50307: gap of 100 bp
* 50308 53363: contig of 3056 bp in length
* 53364 53463: gap of 100 bp
* 53464 56760: contig of 3297 bp in length
* 56761 56860: gap of 100 bp
* 56861 61207: contig of 4347 bp in length
* 61208 61307: gap of 100 bp
* 61308 65984: contig of 4677 bp in length
* 65985 66084: gap of 100 bp
* 66085 72072: contig of 5988 bp in length
* 72073 72172: gap of 100 bp
* 72173 77741: contig of 5569 bp in length
* 77742 77841: gap of 100 bp
* 77842 85850: contig of 8009 bp in length
* 85851 85950: gap of 100 bp
* 85951 92902: contig of 6952 bp in length
* 92903 93002: gap of 100 bp
* 93003 103668: contig of 10666 bp in length
* 103669 103768: gap of 100 bp
* 103769 109322: contig of 5554 bp in length

```
            *   109323 109422: gap of      100 bp
            *   109423   118526: contig of 9104 bp in length
            *   118527 118626: gap of      100 bp
            *   118627   128874: contig of 10248 bp in length
            *   128875 128974: gap of      100 bp
            *   128975   138016: contig of 9042 bp in length
            *   138017 138116: gap of      100 bp
            *   138117   166500: contig of 28384 bp in length
            *   166501 166600: gap of      100 bp
            *   166601   166889: contig of 289 bp in length.
FEATURES             Location/Qualifiers
     source          1. .166889
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-212N14"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .140
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    241. .1566
                     /note="assembly_fragment"
     misc_feature    1667. .26279
                     /note="assembly_fragment"
     misc_feature    26380. .27676
                     /note="assembly_fragment"
     misc_feature    27777. .29820
                     /note="assembly_fragment"
     misc_feature    29921. .33216
                     /note="assembly_fragment"
     misc_feature    33317. .36627
                     /note="assembly_fragment"
     misc_feature    36728. .39382
                     /note="assembly_fragment"
     misc_feature    39483. .42417
                     /note="assembly_fragment"
     misc_feature    42518. .46306
                     /note="assembly_fragment"
     misc_feature    46407. .50207
                     /note="assembly_fragment"
     misc_feature    50308. .53363
                     /note="assembly_fragment"
     misc_feature    53464. .56760
                     /note="assembly_fragment"
     misc_feature    56861. .61207
                     /note="assembly_fragment"
     misc_feature    61308. .65984
                     /note="assembly_fragment"
     misc_feature    66085. .72072
                     /note="assembly_fragment"
     misc_feature    72173. .77741
                     /note="assembly_fragment"
     misc_feature    77842. .85850
                     /note="assembly_fragment"
     misc_feature    85951. .92902
```

```
                     /note="assembly_fragment"
     misc_feature    93003. .103668
                     /note="assembly_fragment"
     misc_feature    103769. .109322
                     /note="assembly_fragment"
     misc_feature    109423. .118526
                     /note="assembly_fragment"
     misc_feature    118627. .128874
                     /note="assembly_fragment"
     misc_feature    128975. .138016
                     /note="assembly_fragment"
     misc_feature    138117. .166500
                     /note="assembly_fragment"
     misc_feature    166601. .166889
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    43782 a  38337 c  39037 g  43225 t   2508 others
ORIGIN


  Query Match             92.3%;  Score 1201.8;  DB 2;  Length 166889;
  Best Local Similarity   97.5%;  Pred. No. 0;
  Matches 1271;  Conservative   10;  Mismatches   15;  Indels    8;  Gaps    6;

Qy      1 gccactccaagctaccatctgagattgtttcctgccctagagtggtaaaggcgtgaggtc 60
          |||||||||||| |||||||||||||||||||||||||||||||||| |||||||||
Db 108011 GCCACTCCAAGC-ACCATCTGAGATTGTTTCCTGCCCTAGAGTGGTAAAGCCGTGAGGTC 108069

Qy     61 cgtctgccctcagctgtgtccccaggcccagggcgtgcctggcaacannagcaggcctct 120
          |||||||||||||||||||||||||||||||||||||||||||     ||||||||||||
Db 108070 CGTCTGCCCTCAGCTGTGTCCCCAGGCCCAGGGCGTGCCTGGCA--CAGAGCAGGCCTCT 108127

Qy    121 gagaaccagcctcccacgtgagttcatgatagnaagacagcccctcgttcccattcagtg 180
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 108128 GAGAACCAGCCTCCCACGTGAGTTCATGATAGCAAGACAGCCCCTCGTTCCCATTCAGTG 108187

Qy    181 gttggttctgttctttycctggcmataagctccactctg-ymrtcagccamacatttatt 239
          |||||||||||||||:||||||:||| |||||||||||| :::|||||||:|||||||||
Db 108188 GTTGGTTCTGTTCTTTCCCTGGCCATAGGCTCCACTCTGTCAGTCAGCCACACATTTATT 108247

Qy    240 gagtaccagttgttggcaaagcactgttgggcatgaaaagcattaacccagtgaatgagg 299
          |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db 108248 GAGTACCAGTTGTGTGCAAAGCACTGTTGGGCATGAAAAGCATTAACCCAGTGAATGAGG 108307

Qy    300 aggagcttgggtt-gggacggagccmcaraawtacatggcagaccagaaggaaatcagct 358
          |||||||||||||| ||||||||||:||:||:|||||||||||||||||| ||||||||
Db 108308 AGGAGCTTGGGTTGGGGACGGAGCCCCAGAATTACATGGCAGACCAGAAGGGAATCAGCT 108367

Qy    359 caagtagaaaracacgcatgggctcgtgggcgacgcagtgtgtgctgtgtcatctggggc 418
          ||      ||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db 108368 CA--GTAGAAGACACGCATGGGCTCGTGGGCGACGCAGTGTGTGCTGTGTCATCTGGGGC 108425

Qy    419 tgggaggaagtgtcctggatcaggagttccaggagcccaggaggagtggacgggtcagtg 478
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db 108426 TGGGAGGAAGTGTCCTGGGTCAGGAGTTCCAGGAGCCCAGGAGGAGTGGACGGGTCAGTG 108485
```

```
Qy    479 cagagccagcccgcaatcaggggaagaaaacacggccaaggccaggccttcacggggagc 538
          ||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||
Db 108486 CAGAGCCAGCCCGCATTCAGGGG-AGAAAACACGGCCAAGGCCAGGCCTTCACGGGGAGC 108544

Qy    539 ccagcgtgggctgcacatctgcactctccaggctagttttggtgcccacatgctctgcag 598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108545 CCAGCGTGGGCTGCACATCTGCACTCTCCAGGCTAGTTTTGGTGCCCACATGCTCTGCAG 108604

Qy    599 ggtctgggcactgtggcagcggcagcaggcttccctgttgctagtccagctgctgaaact 658
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108605 GGTCTGGGCACTGTGGCAGCGGCAGCAGGCTTCCCTGTTGCTAGTCCAGCTGCTGAAACT 108664

Qy    659 ccagggagagtcaaaaagttcccaaatacagaggcgtggctggtagtccttcccgggaat 718
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108665 CCAGGGAGAGTCAAAAAGTTCCCAAATACAGAGGCGTGGCTGGTAGTCCTTCCCGGGAAT 108724

Qy    719 tcttcttgcttcccgctttctgtggaactctgccttccccactctgcctctctgcttgtt 778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108725 TCTTCTTGCTTCCCGCTTTCTGTGGAACTCTGCCTTCCCCACTCTGCCTCTCTGCTTGTT 108784

Qy    779 cctgggccccaggacctctttcccatcttcgatctcttaagtcataccttgggaggcctc 838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108785 CCTGGGCCCCAGGACCTCTTTCCCATCTTCGATCTCTTAAGTCATACCTTGGGAGGCCTC 108844

Qy    839 ccccagcccgccgtgtaaagagggctgtcacagcttctgctgtcacagaagcattacaat 898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108845 CCCCAGCCCGCCGTGTAAAGAGGGCTGTCACAGCTTCTGCTGTCACAGAAGCATTACAAT 108904

Qy    899 gtgcaggtgcctgttaacatctgccttccccactgatctggagctccacaagggagaggg 958
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108905 GTGCAGGTGCCTGTTAACATCTGCCTTCCCCACTGATCTGGAGCTCCACAAGGGAGAGGG 108964

Qy    959 cacacccagtaggtatgtgtgggatggataggagggtggatgacacccagtagatgtgta 1018
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 108965 CACACCCAGTAGGTATGTGTGGGATGGATAGGAGGGTGGATGACACCCAGTAGATGTGTA 109024

Qy   1019 tgggatggataggagggtggatgacacccagtaggtgtgtatgggatggatgggagggtg 1078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109025 TGGGATGGATAGGAGGGTGGATGACACCCAGTAGGTGTGTATGGGATGGATGGGAGGGTG 109084

Qy   1079 ggtgacccctagtagatgtgggggggggtgggtgggtgacccccagtaggtgtgtgtggca 1138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109085 GGTGACCCCTAGTAGATGTGGGGGGGGGTGGGTGGGTGACCCCCAGTAGGTGTGTGTGGCA 109144

Qy   1139 tggataggtgacccccagtagacgtttgtgggacggatgggagggtaggtaagtgacccc 1198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109145 TGGATAGGTGACCCCCAGTAGACGTTTGTGGGACGGATGGGAGGGTAGGTAAGTGACCCC 109204

Qy   1199 caggaggcgtctatagggcaggtgggtggatgtggatgaacagcaccttgtttcttcttc 1258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 109205 CAGGAGGCGTCTATAGGGCAGGTGGGTGGATGTGGATGAACAGCACCTTGTTTCTTCTTC 109264

Qy   1259 ccaggtggcttctggcacagcagcttaattgaccggaacctcat 1302
          |||||||||||||||||||||||||||||||||||||||||||
Db 109265 CCAGGTGGCTTCTGGCACAGCAGCTTAATTGACCGGAACCTCAT 109308
```

RESULT 2
AL158207/c
LOCUS AL158207 169963 bp DNA linear PRI 25-SEP-2001
DEFINITION Human DNA sequence from clone RP11-409K20 on chromosome 9 Contains the TOR1B gene for torsin family 1 member B (torsin B) (DQ1), the DYT1 gene for 'dystonia 1, torsion' (autosomal dominant; torsin A) (DQ2, TOR1A), the gene for hepatocellular carcinoma-associated antigen 59 (HSPC220, LOC51759), the USP20 gene for ubiquitin specific protease 20 (KIAA1003), and the gene for formin-binding protein 17 (FBP17, includes KIAA0554, FLJ13619, FLJ10754 and FLJ10113). Contains ESTs, STSs, GSSs and four CpG islands, complete sequence.
ACCESSION AL158207
VERSION AL158207.15 GI:12717949
KEYWORDS HTG; CpG island; DQ1; dystonia; DYT1; FBP17; FLJ10113; FLJ10754; FLJ13619; formin-binding; hepatocellular; HSPC220; KIAA0554; KIAA1003; LOC51759; protease; TOR1B; torsin; ubiquitin; USP20.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169963)
AUTHORS Babbage,A.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Feb 8, 2001 this sequence version replaced gi:12657099.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-409K20 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-409K20 The true left end of clone RP11-138E2 is at 118932 in this sequence.
FEATURES Location/Qualifiers

```
source          1. .169963
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="9"
                /clone="RP11-409K20"
                /clone_lib="RPCI-11.2"
repeat_region   5. .86
                /note="MSTC repeat: matches 46. .126 of consensus"
misc_feature    28. .462
                /note="match: GSS: Em:AQ718881"
repeat_region   817. .992
                /note="Charlie2 repeat: matches 7. .195 of consensus"
misc_feature    complement(2510. .2941)
                /note="match: GSS: Em:AQ041615"
misc_feature    2944. .3096
                /note="match: GSS: Em:B74700"
misc_feature    3329. .4807
                /note="CpG island"
                /evidence=not_experimental
mRNA            join(4205. .4464,5126. .5391,8241. .8416,9958. .10085,
                10395. .12334)
                /gene="TOR1B"
                /note="match: cDNAs: Em:AF007872 Em:AJ297743
                match: ESTs: Em:AI815528 Em:AW160403 Em:AW972065
                Em:BF313148 Em:AA112625 Em:BE740991 Em:BE563034
                Em:BE893335 Em:AV728123 Em:AW952051 Em:AW148938
                Em:BE502754 Em:AW016676 Em:AI223067 Em:BE108689
                Em:AI808893 Em:AW173267 Em:AI185247"
                /product="bA409K20.1.1 (torsin family 1, member B (torsin
                B) (DQ1))"
                /evidence=not_experimental
gene            4205. .12334
                /gene="TOR1B"
CDS             join(4266. .4464,5126. .5391,8241. .8416,9958. .10085,
                10395. .10636)
                /gene="TOR1B"
                /note="match: proteins: Tr:O14657"
                /codon_start=1
                /evidence=not_experimental
                /product="bA409K20.1.1 (torsin family 1, member B (torsin
                B) (DQ1))"
                /protein_id="CAC88165.1"
                /db_xref="GI:15787707"
                /translation="MLRAGWLRGAAALALLLAARVVAAFEPITVGLAIGAASAITGYL
                SYNDIYCRFAECCREERPLNASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKP
                LTLSLHGWAGTGKNFVSQIVAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQK
                WIRGNVSACANSVFIFDEMDKLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLSNAGG
                DLITKTALDFWRAGRKREDIQLKDLEPVLSVGVFNNKHSGLWHSGLIDKNLIDYFIPF
                LPLEYRHVKMCVRAEMRARGSAIDEDIVTRVAEEMTFFPRDEKIYSDKGCKTVQSRLD
                FH"
mRNA            join(4321. .4464,5126. .5391,11280. .11571)
                /gene="TOR1B"
                /note="isoform 3
                match: ESTs: Em:BF058863 Em:BE315222"
                /product="bA409K20.1.3 (torsin family 1, member B (torsin
                B) (DQ1), putative isoform 3)"
                /evidence=not_experimental
```

```
CDS             join(<4321. .4464,5126. .5391,11280. .11294)
                /gene="TOR1B"
                /codon_start=3
                /evidence=not_experimental
                /product="bA409K20.1.3 (torsin family 1, member B (torsin
                B) (DQ1), putative isoform 3)"
                /protein_id="CAC88166.1"
                /db_xref="GI:15787708"
                /translation="RVVAAFEPITVGLAIGAASAITGYLSYNDIYCRFAECCREERPL
                NASALKLDLEEKLFGQHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQI
                VAENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQSSLT"
mRNA            join(5126. .5266,8241. .8416,9958. .10085,10395. .10456)
                /gene="TOR1B"
                /note="match: ESTs: Em:AI468027"
                /product="bA409K20.1.2 (isoform 2)"
                /evidence=not_experimental
mRNA            join(5159. .5391,8241. .8416,11280. .11319)
                /gene="TOR1B"
                /note="isoform 4
                match: ESTs: Em:AI568476"
                /product="bA409K20.1.4 (torsin family 1, member B (torsin
                B) (DQ1), putative isoform 4)"
                /evidence=not_experimental
CDS             join(<5159. .5391,8241. .8416,11280. .11289)
                /gene="TOR1B"
                /codon_start=3
                /evidence=not_experimental
                /product="bA409K20.1.4 (torsin family 1, member B (torsin
                B) (DQ1), putative isoform 4)"
                /protein_id="CAC88167.1"
                /db_xref="GI:15787709"
                /translation="QHLATEVIFKALTGFRNNKNPKKPLTLSLHGWAGTGKNFVSQIV
                AENLHPKGLKSNFVHLFVSTLHFPHEQKIKLYQDQLQKWIRGNVSACANSVFIFDEMD
                KLHPGIIDAIKPFLDYYEQVDGVSYRKAIFIFLRVH"
misc_feature    5188. .5526
                /gene="TOR1B"
                /note="match: STS: Em:G24606"
repeat_region   7370. .7432
                /note="MER61E repeat: matches 128. .190 of consensus"
misc_feature    complement(11923. .12334)
                /note="match: STS: Em:G27406"
misc_feature    complement(12097. .12334)
                /note="match: STS: Em:G24725"
polyA_signal    12313. .12318
                /gene="TOR1B"
polyA_site      12334
                /gene="TOR1B"
polyA_site      complement(13997)
                /gene="DYT1"
mRNA            complement(join(13997. .15275,19573. .19700,19798. .19973,
                23634. .23899,24961. .25180))
                /gene="DYT1"
                /note="match: cDNAs: Em:AF007871
                match: ESTs: Em:BE272533 Em:BE314317 Em:BE784377
                Em:BF203163 Em:BE622540 Em:AI039978 Em:AI770117
                Em:AW050630 Em:AI970719 Em:BE463967 Em:AI374678
                Em:AI167967 Em:AI127274 Em:AI699731 Em:AW001722
```

```
                     Em:AI301894 Em:AW080988"
                     /product="bA409K20.2 (dystonia 1, torsion (autosomal
                     dominant; torsin A) (DQ2, TOR1A))"
                     /evidence=not_experimental
     gene            complement(13997. .25180)
                     /gene="DYT1"
     polyA_signal    complement(14010. .14015)
                     /gene="DYT1"
     misc_feature    14016. .14298
                     /note="match: STS: Em:G30092"
     misc_feature    complement(14429. .14885)
                     /gene="DYT1"
                     /note="match: GSS: Em:B69651"
     misc_feature    complement(14469. .14876)
                     /gene="DYT1"
                     /note="match: GSS: Em:B48142"
     misc_feature    complement(14494. .14860)
                     /gene="DYT1"
                     /note="match: GSS: Em:AQ566167"
     polyA_site      complement(14632)
                     /gene="DYT1"
     misc_feature    14650. .15099
                     /note="match: STS: Em:G60041 Em:G60042"
     misc_feature    14807. .14914
                     /note="match: STS: Em:G43378 Em:G43379"
     misc_feature    14885. .15212
                     /note="match: GSS: Em:AQ213491"
     misc_feature    14890. .15392
                     /note="match: GSS: Em:AQ482600"
     CDS             complement(join(15025. .15275,19573. .19700,19798. .19973,
                     23634. .23899,24961. .25138))

  Query Match             92.3%;  Score 1201.8;  DB 9;  Length 169963;
  Best Local Similarity   97.5%;  Pred. No. 0;
  Matches 1271;  Conservative   10;  Mismatches   15;  Indels    8;  Gaps    6;

Qy       1 gccactccaagctaccatctgagattgtttcctgccctagagtggtaaaggcgtgaggtc 60
           |||||||||||| |||||||||||||||||||||||||||||||||| |||||||||
Db   16533 GCCACTCCAAGC-ACCATCTGAGATTGTTTCCTGCCCTAGAGTGGTAAAGCCGTGAGGTC 16475

Qy      61 cgtctgccctcagctgtgtccccaggcccagggcgtgcctggcaacannagcaggcctct 120
           ||||||||||||||||||||||||||||||||||||||||||||     ||||||||||||
Db   16474 CGTCTGCCCTCAGCTGTGTCCCCAGGCCCAGGGCGTGCCTGGCA--CAGAGCAGGCCTCT 16417

Qy     121 gagaaccagcctcccacgtgagttcatgatagnaagacagcccctcgttcccattcagtg 180
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   16416 GAGAACCAGCCTCCCACGTGAGTTCATGATAGCAAGACAGCCCCTCGTTCCCATTCAGTG 16357

Qy     181 gttggttctgttctttycctggcmataagctccactctg-ymrtcagccamacatttatt 239
           |||||||||||||||:||||||:||| |||||||||||| :::|||||||:|||||||||
Db   16356 GTTGGTTCTGTTCTTTCCCTGGCCATAGGCTCCACTCTGTCAGTCAGCCACACATTTATT 16297

Qy     240 gagtaccagttgttggcaaagcactgttgggcatgaaaagcattaacccagtgaatgagg 299
           |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db   16296 GAGTACCAGTTGTGTGCAAAGCACTGTTGGGCATGAAAAGCATTAACCCAGTGAATGAGG 16237

Qy     300 aggagcttgggtt-gggacggagccmcaraawtacatggcagaccagaaggaaatcagct 358
```

```
          |||||||||||||  ||||||||||:||:||:||||||||||||||||||  ||||||||
Db  16236 AGGAGCTTGGGTTGGGGACGGAGCCCCAGAATTACATGGCAGACCAGAAGGGAATCAGCT 16177

Qy    359 caagtagaaaaracacgcatgggctcgtgggcgacgcagtgtgtgctgtgtcatctggggc 418
          ||      ||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16176 CA--GTAGAAGACACGCATGGGCTCGTGGGCGACGCAGTGTGTGCTGTGTCATCTGGGGC 16119

Qy    419 tgggaggaagtgtcctggatcaggagttccaggagcccaggaggagtggacgggtcagtg 478
          |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db  16118 TGGGAGGAAGTGTCCTGGGTCAGGAGTTCCAGGAGCCCAGGAGGAGTGGACGGGTCAGTG 16059

Qy    479 cagagccagcccgcaatcaggggaagaaaacacggccaaggccaggccttcacggggagc 538
          |||||||||||||||  |||||||  ||||||||||||||||||||||||||||||||||
Db  16058 CAGAGCCAGCCCGCATTCAGGGG-AGAAAACACGGCCAAGGCCAGGCCTTCACGGGGAGC 16000

Qy    539 ccagcgtgggctgcacatctgcactctccaggctagttttggtgcccacatgctctgcag 598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15999 CCAGCGTGGGCTGCACATCTGCACTCTCCAGGCTAGTTTTGGTGCCCACATGCTCTGCAG 15940

Qy    599 ggtctgggcactgtggcagcggcagcaggcttccctgttgctagtccagctgctgaaact 658
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15939 GGTCTGGGCACTGTGGCAGCGGCAGCAGGCTTCCCTGTTGCTAGTCCAGCTGCTGAAACT 15880

Qy    659 ccagggagagtcaaaaagttcccaaatacagaggcgtggctggtagtccttcccgggaat 718
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15879 CCAGGGAGAGTCAAAAAGTTCCCAAATACAGAGGCGTGGCTGGTAGTCCTTCCCGGGAAT 15820

Qy    719 tcttcttgcttcccgctttctgtggaactctgccttccccactctgcctctctgcttgtt 778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15819 TCTTCTTGCTTCCCGCTTTCTGTGGAACTCTGCCTTCCCCACTCTGCCTCTCTGCTTGTT 15760

Qy    779 cctgggccccaggacctctttcccatcttcgatctcttaagtcataccttgggaggcctc 838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15759 CCTGGGCCCCAGGACCTCTTTCCCATCTTCGATCTCTTAAGTCATACCTTGGGAGGCCTC 15700

Qy    839 ccccagcccgccgtgtaaagagggctgtcacagcttctgctgtcacagaagcattacaat 898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15699 CCCCAGCCCGCCGTGTAAAGAGGGCTGTCACAGCTTCTGCTGTCACAGAAGCATTACAAT 15640

Qy    899 gtgcaggtgcctgttaacatctgccttccccactgatctggagctccacaagggagaggg 958
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15639 GTGCAGGTGCCTGTTAACATCTGCCTTCCCCACTGATCTGGAGCTCCACAAGGGAGAGGG 15580

Qy    959 cacacccagtaggtatgtgtgggatggataggagggtggatgacacccagtagatgtgta 1018
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15579 CACACCCAGTAGGTATGTGTGGGATGGATAGGAGGGTGGATGACACCCAGTAGATGTGTA 15520

Qy   1019 tgggatggataggagggtggatgacacccagtaggtgtgtatgggatggatgggagggtg 1078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15519 TGGGATGGATAGGAGGGTGGATGACACCCAGTAGGTGTGTATGGGATGGATGGGAGGGTG 15460

Qy   1079 ggtgacccctagtagatgtgggggggggtgggtgggtgacccccagtaggtgtgtgtggca 1138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15459 GGTGACCCCTAGTAGATGTGGGGGGGGTGGGTGGGTGACCCCCAGTAGGTGTGTGTGGCA 15400

Qy   1139 tggataggtgacccccagtagacgtttgtgggacggatgggagggtaggtaagtgacccc 1198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  15399 TGGATAGGTGACCCCCAGTAGACGTTTGTGGGACGGATGGGAGGGTAGGTAAGTGACCCC 15340

Qy   1199 caggaggcgtctatagggcaggtgggtggatgtggatgaacagcaccttgtttcttcttc 1258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  15339 CAGGAGGCGTCTATAGGGCAGGTGGGTGGATGTGGATGAACAGCACCTTGTTTCTTCTTC 15280

Qy   1259 ccaggtggcttctggcacagcagcttaattgaccggaacctcat 1302
          |||||||||||||||||||||||||||||||||||||||||||
Db  15279 CCAGGTGGCTTCTGGCACAGCAGCTTAATTGACCGGAACCTCAT 15236
```

```
=> fil reg; d que l6
FILE 'REGISTRY' ENTERED AT 10:54:18 ON 07 JUN 2002
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2002 American Chemical Society (ACS)

STRUCTURE FILE UPDATES:    5 JUN 2002  HIGHEST RN 426206-38-4
DICTIONARY FILE UPDATES:   5 JUN 2002  HIGHEST RN 426206-38-4

TSCA INFORMATION NOW CURRENT THROUGH January 7, 2002

  Please note that search-term pricing does apply when
  conducting SmartSELECT searches.

Crossover limits have been increased.  See HELP CROSSOVER for details.

Calculated physical property data is now available.  See HELP PROPERTIES
for more information.  See STNote 27, Searching Properties in the CAS
Registry File, for complete details:
http://www.cas.org/ONLINE/STN/STNOTES/stnotes27.pdf


L4            63 SEA FILE=REGISTRY ABB=ON  GCAAAACAGGGCUUUGUACCG|CGGUACAAAGCCCUG
                 UUUUGC|AGUAGAGACGCGGGUAGAUG|CAUCUACCCGCGUCUCUACU|GCGUCUCUACUGCC
                 UCUUCG|CGAAGAGGCAGUAGAGACGC|AUGCCCUGGUCCUAGUUCAG|UACGGGACCAGGAU
                 CAAGUC|GGUUUCGCAAGGUGCUUGGA|UCCAAGCACCUUGCGAAACC/SQSN
L5           109 SEA FILE=REGISTRY ABB=ON  GGGAUUCCAAACUUCCAUCC|GGAUGGAAGUUUGGAA
                 UCCC|UCCAUGGGGUUGGUAGGAAC|GUUCCUACCAACCCCAUGGA|GGUGACAGAGUAAAAC
                 UAUCUG|CAGAUAGUUUUACUCUGUCACC|GACCCCCAGUAGACGUUUGU|ACAAACGUCUAC
                 UGGGGGUC|GUAAAAAAUCAUGAGCCCUGC|GCAGGGCUCAUGAUUUUUUAC/SQSN
L6            10 SEA FILE=REGISTRY ABB=ON  (L4 OR L5) AND SQL<101


=> d rn cn kwic nte l6 1-10; fil capl; s l6

L6   ANSWER 1 OF 10  REGISTRY  COPYRIGHT 2002 ACS
RN   367568-27-2  REGISTRY
CN   DNA, d(G-T-A-A-A-A-A-A-T-C-A-T-G-A-G-C-C-C-T-G-C)  (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   33: PN: US20010029015 SEQID: 39 claimed DNA
SQL  21

SEQ      1 gtaaaaaatc atgagccctg c
           ========== ========== =
HITS AT:   1-21

L6   ANSWER 2 OF 10  REGISTRY  COPYRIGHT 2002 ACS
RN   367568-26-1  REGISTRY
CN   DNA, d(G-A-C-C-C-C-C-A-G-T-A-G-A-C-G-T-T-T-G-T)  (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   32: PN: US20010029015 SEQID: 38 claimed DNA
SQL  20

SEQ      1 gacccccagt agacgtttgt
           ========== ==========
HITS AT:   1-20

L6   ANSWER 3 OF 10  REGISTRY  COPYRIGHT 2002 ACS
RN   367568-25-0  REGISTRY
CN   DNA, d(G-G-T-G-A-C-A-G-A-G-T-A-A-A-A-C-T-A-T-C-T-G)  (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   31: PN: US20010029015 SEQID: 37 claimed DNA
SQL  22
```

SEQ 1 ggtgacagag taaaactatc tg
========== ========== ==
HITS AT: 1-22

L6 ANSWER 4 OF 10 REGISTRY COPYRIGHT 2002 ACS
RN 367568-24-9 REGISTRY
CN DNA, d(T-C-C-A-T-G-G-G-G-T-T-G-G-T-A-G-G-A-A-C) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 30: PN: US20010029015 SEQID: 36 claimed DNA
SQL **20**

SEQ 1 tccatggggt tggtaggaac
========== ==========
HITS AT: 1-20

L6 ANSWER 5 OF 10 REGISTRY COPYRIGHT 2002 ACS
RN 367568-23-8 REGISTRY
CN DNA, d(G-G-G-A-T-T-C-C-A-A-A-C-T-T-C-C-A-T-C-C) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 29: PN: US20010029015 SEQID: 35 claimed DNA
SQL **20**

SEQ 1 gggattccaa acttccatcc
========== ==========
HITS AT: 1-20

L6 ANSWER 6 OF 10 REGISTRY COPYRIGHT 2002 ACS
RN 367568-22-7 REGISTRY
CN DNA, d(G-G-T-T-T-C-G-C-A-A-G-G-T-G-C-T-T-G-G-A) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 28: PN: US20010029015 SEQID: 34 claimed DNA
SQL **20**

SEQ 1 ggtttcgcaa ggtgcttgga
========== ==========
HITS AT: 1-20

L6 ANSWER 7 OF 10 REGISTRY COPYRIGHT 2002 ACS
RN 367568-21-6 REGISTRY
CN DNA, d(A-T-G-C-C-C-T-G-G-T-C-C-T-A-G-T-T-C-A-G) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 27: PN: US20010029015 SEQID: 33 claimed DNA
SQL **20**

SEQ 1 atgccctggt cctagttcag
========== ==========
HITS AT: 1-20

L6 ANSWER 8 OF 10 REGISTRY COPYRIGHT 2002 ACS
RN 367568-20-5 REGISTRY
CN DNA, d(G-C-G-T-C-T-C-T-A-C-T-G-C-C-T-C-T-T-C-G) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 26: PN: US20010029015 SEQID: 32 claimed DNA
SQL **20**

SEQ 1 gcgtctctac tgcctcttcg
========== ==========
HITS AT: 1-20

L6 ANSWER 9 OF 10 REGISTRY COPYRIGHT 2002 ACS
RN 367568-19-2 REGISTRY
CN DNA, d(A-G-T-A-G-A-G-A-C-G-C-G-G-G-T-A-G-A-T-G) (9CI) (CA INDEX NAME)

OTHER NAMES:
CN 25: PN: US20010029015 SEQID: 31 claimed DNA
SQL **20**

SEQ 1 agtagagacg cgggtagatg
========== ==========
HITS AT: 1-20

L6 ANSWER 10 OF 10 REGISTRY COPYRIGHT 2002 ACS
RN 367568-18-1 REGISTRY
CN DNA, d(G-C-A-A-A-A-C-A-G-G-G-C-T-T-T-G-T-A-C-C-G) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 24: PN: US20010029015 SEQID: 30 claimed DNA
SQL **21**

SEQ 1 gcaaaacagg gctttgtacc g
========== ========== =
HITS AT: 1-21

FILE 'CAPLUS' ENTERED AT 10:54:35 ON 07 JUN 2002
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2002 AMERICAN CHEMICAL SOCIETY (ACS)

FILE COVERS 1907 - 7 Jun 2002 VOL 136 ISS 23
FILE LAST UPDATED: 5 Jun 2002 (20020605/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

CAS roles have been modified effective December 16, 2001. Please check your SDI profiles to see if they need to be revised. For information on CAS roles, enter HELP ROLES at an arrow prompt or use the CAS Roles thesaurus (/RL field) in this file.

L7 1 L6

=> d ibib ab hitrn

L7 ANSWER 1 OF 1 CAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2001:748274 CAPLUS
DOCUMENT NUMBER: 135:316961
TITLE: Nucleic acid sequences for torsins encoded by human genes DYT1/TOR1A, TOR1B, and torsin-related genes and their use in detecting torsion dystonia or neuronal disease
INVENTOR(S): Ozelius, Laurie J.; Breakefield, Xandra O.
PATENT ASSIGNEE(S): The General Hospital Corp., USA
SOURCE: U.S. Pat. Appl. Publ., 85 pp., Cont.-in-part of U. S.

Ser. No. 461,921, abandoned.
CODEN: USXXCO
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 3
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 2001029015	A1	20011011	US 2001-772105	20010126
US 6387616	B1	20020514	US 1998-218363	19981222

PRIORITY APPLN. INFO.: US 1997-50244P P 19970619
US 1998-99454 A2 19980618
US 1998-218363 A2 19981222
US 1999-461921 B2 19991215

AB The present invention relates to methods of detecting mutations and polymorphisms in the torsin gene, torsin-related genes, methods of detecting neuronal diseases mediated by these mutations and polymorphisms and nucleic acids used in these methods. A CAG deletion in exon 5 of the human gene DYT1/TOR1A and the DQ2 cDNA of this gene (encoding torsinA) causes early onset dystonia. The exon/intron structure and cDNAs of gene DYT1 have been characterized by sequence anal. and genetic polymorphisms have been identified. An adjacent gene on human chromosome 9q34, named TOR1B, encodes a homologous protein torsinB. Homol. searches have identified human and mouse cDNAs for torsin-related proteins encoded by genes TORP1 and TORP2. This invention provides for further anal. of the torsinA gene family and its role in human disease.

IT **367568-18-1 367568-19-2 367568-20-5**
367568-21-6 367568-22-7 367568-23-8
367568-24-9 367568-25-0 367568-26-1
367568-27-2
RL: ARG (Analytical reagent use); THU (Therapeutic use); ANST (Analytical study); BIOL (Biological study); USES (Uses)
(human gene DYT1/TOR1A specific primer; nucleic acid sequences for torsins encoded by human genes TOR1A(DYT1), TOR1B, and torsin-related genes and their use in detecting torsion dystonia or neuronal disease)

=> fil hom
FILE 'HOME' ENTERED AT 10:54:57 ON 07 JUN 2002

SID 39 100%

RESULT 1
AAC69659/c
ID AAC69659 standard; cDNA; 853 BP.
XX
AC AAC69659;
XX
DT 30-JAN-2001 (first entry)
XX
DE Human torsin A coding sequence.
XX
KW Cytostatic; vaccine; human; breast tumour; antigen; breast cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200052165-A2.
XX
PD 08-SEP-2000.
XX
PF 29-FEB-2000; 2000WO-US05431.
XX
PR 04-MAR-1999; 99US-0262505.
PR 19-MAR-1999; 99US-0272886.
PR 17-SEP-1999; 99US-0396313.
XX
PA (CORI-) CORIXA CORP.
XX
PI Lodes MJ;
XX
DR WPI; 2000-572184/53.
XX
PT Breast tumor antigen polypeptides and polynucleotides, useful for
PT manufacturing vaccines and compositions for treating, diagnosing, and
PT monitoring breast cancer -
XX
PS Claim 16; Fig 1; 140pp; English.
XX
CC The present invention relates to immunogenic portions of new human
CC breast tumour antigens (AAB28183-B28214) and their coding sequences
CC (AAC69645-C69804). The breast tumour antigen polypeptides of the present
CC invention and their coding sequences are useful for inhibiting the
CC development of breast cancer in a patient. The breast tumour antigen
CC polypeptides and polynucleotides may be used in vaccines and
CC pharmaceutical compositions for treating breast cancer, and for
CC diagnosing and monitoring the cancer. The present sequence is a coding
CC sequence for the immunogenic portion for one such human breast cancer
CC tumour antigen.
XX
SQ Sequence 853 BP; 233 A; 177 C; 187 G; 256 T; 0 other;

39 only

102(a) type if no prior

Query Match 100.0%; Score 21; DB 21; Length 853;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaaaaaatcatgagccctgc 21
||||||||||||||||||||
Db 85 GTAAAAAATCATGAGCCCTGC 65

type
102 (b) misla
102 (a) priority

31, 32, 39

RESULT 2
AAV99925/c
ID AAV99925 standard; cDNA; 2072 BP.
XX
AC AAV99925;
XX
DT 12-MAY-1999 (first entry)
XX
DE DYT1 torsion dystonia gene (torsinA).
XX
KW Torsion dystonia; DYT1; torsinA; torsinB; DQ2; DQ1;
KW neurotransmission; movement disorder; chorea; tremor; rigidity;
KW Huntingtons disease; Parkinsons disease; diagnosis; prognosis;
KW prevention; treatment; neurology; neuropathology; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 43..1041
FT /*tag= a
FT /product= TorsinA_protein
XX
PN WO9857984-A2.
XX
PD 23-DEC-1998.
XX
PF 19-JUN-1998; 98WO-US12776.
XX
PR 18-JUN-1998; 98US-0099454.
PR 19-JUN-1997; 97US-0050244.
XX
PA (BREA/) BREAKEFIELD X.
PA (OZEL/) OZELIUS L J.
XX
PI Breakefield X, Ozelius LJ;
XX
DR WPI; 1999-080947/07.
DR P-PSDB; AAW81057.
XX
PT New isolated torsion dystonia genes - used to develop products for
PT the diagnosis, prognosis, prevention and treatment of torsion
PT dystonia
XX
PS Example 2; Page 106-109; 138pp; English.
XX
CC Movement disorders generally comprise some kind of aberrant
CC neurotransmission. These often manifest themselves as
CC uncontrollable body movements such as chorea in Huntington's
CC disease, tremor and rigidity in Parkinson's disease and twisting
CC contractions in torsion dystonia. Dystonic syptoms can be
CC secondary to neurological conditions but primary or torsion
CC dystonia is characterised by a lack of other neurologic involvement
CC and the absence of any distinct neuropathology. Clinical
CC manifestations of torsion dystonia can affect many different body
CC regions. Novel torsion dystonia genes, their polypeptide and

CC protein products, recombinant nucleic acids comprising them, cells
CC transformed by them or recombinant molecules in which they are
CC contained, as well as antibody molecules directed against them,
CC can be used to develop products for the diagnosis, prognosis,
CC prevention and treatment of torsion dystonia. In particular, the
CC torsin polypeptides can be used to treat torsion dystonia. This
CC sequence is a composite nucleotide sequence of the torsinA gene.
XX
SQ Sequence 2072 BP; 530 A; 489 C; 510 G; 543 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 2072;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaaaaaatcatgagccctgc 21
||||||||||||||||||||
Db 1315 GTAAAAAATCATGAGCCCTGC 1295

162-181 - SID 32
172-153 has 31

RESULT 3
AAV59658/c
ID AAV59658 standard; DNA; 2117 BP. Est - no rej.
XX
AC AAV59658;
XX
DT 19-JAN-1999 (first entry)
XX
DE Human secreted protein gene 148 clone HSKGO26.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX
OS Homo sapiens.
XX
PN WO9839448-A2.
XX
PD 11-SEP-1998.
XX
PF 06-MAR-1998; 98WO-US04493.
XX
PR 02-OCT-1997; 97US-0061060.
PR 07-MAR-1997; 97US-0038621.
PR 07-MAR-1997; 97US-0040161.
PR 07-MAR-1997; 97US-0040162.
PR 07-MAR-1997; 97US-0040163.
PR 07-MAR-1997; 97US-0040333.
PR 07-MAR-1997; 97US-0040334.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.

31, 32, 33

102(b)

PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043568.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.

PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057669.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
XX
DR WPI; 1998-506364/43.
DR P-PSDB; AAW74876.
XX
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
PS Claim 1; Page 383-384; 721pp; English.
XX
CC This sequence represents a nucleic acid molecule designated Gene 148
CC from the human cDNA clone HSKGO26 (deposited as clone ATCC 97903 and
CC ATCC 209049) which encodes a secreted human protein. The gene can be
CC used to generate fusion proteins by linking to the gene to a human
CC immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
CC the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)

CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAV59511 for described uses).
XX
SQ Sequence 2117 BP; 556 A; 495 C; 516 G; 547 T; 3 other;

Query Match 100.0%; Score 21; DB 19; Length 2117;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaaaaaatcatgagccctgc 21
||||||||||||||||||||
Db 1304 GTAAAAAATCATGAGCCCTGC 1284

191-170 - SID 32
161-142 - SID 31

RESULT 4
AAV99923/c
30,31, ID AAV99923 standard; DNA; 2597 BP.
XX
32, 39 AC AAV99923;
XX
DT 12-MAY-1999 (first entry)
XX
DE DYT1 torsion dystonia gene (torsinA, clone DQ2).
XX
KW Torsion dystonia; DYT1; torsinA; torsinB; DQ2; DQ1;
KW neurotransmission; movement disorder; chorea; tremor; rigidity;
KW Huntingtons disease; Parkinsons disease; diagnosis; prognosis;
KW prevention; treatment; neurology; neuropathology; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 568..1566
FT /*tag= a
FT /product= TorsinA_protein
XX
PN WO9857984-A2.
XX
PD 23-DEC-1998.
XX
PF 19-JUN-1998; 98WO-US12776.
XX
PR 18-JUN-1998; 98US-0099454.
PR 19-JUN-1997; 97US-0050244.
XX
PA (BREA/) BREAKEFIELD X.
PA (OZEL/) OZELIUS L J.
XX
PI Breakefield X, Ozelius LJ;
XX
DR WPI; 1999-080947/07.

DR P-PSDB; AAW81055.
XX
PT New isolated torsion dystonia genes - used to develop products for
PT the diagnosis, prognosis, prevention and treatment of torsion
PT dystonia
XX
PS Claim 2; Page 94-97; 138pp; English.
XX
CC Movement disorders generally comprise some kind of aberrant
CC neurotransmission. These often manifest themselves as
CC uncontrollable body movements such as chorea in Huntington's
CC disease, tremor and rigidity in Parkinson's disease and twisting
CC contractions in torsion dystonia. Dystonic syptoms can be
CC secondary to neurological conditions but primary or torsion
CC dystonia is characterised by a lack of other neurologic involvement
CC and the absence of any distinct neuropathology. Clinical
CC manifestations of torsion dystonia can affect many different body
CC regions. Novel torsion dystonia genes, their polypeptide and
CC protein products, recombinant nucleic acids comprising them, cells
CC transformed by them or recombinant molecules in which they are
CC contained, as well as antibody molecules directed against them,
CC can be used to develop products for the diagnosis, prognosis,
CC prevention and treatment of torsion dystonia. In particular, the
CC torsin polypeptides can be used to treat torsion dystonia. This
CC sequence encodes the torsion dystonia protein TorsinA and was
CC isolated from human adult substantia nigra, hippocampus and
CC frontal cortex.
XX
SQ Sequence 2597 BP; 652 A; 623 C; 656 G; 658 T; 8 other;

Query Match 100.0%; Score 21; DB 20; Length 2597;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gtaaaaaatcatgagccctgc 21
||||||||||||||||||||
Db 1840 GTAAAAAATCATGAGCCCTGC 1820

687 - 706 SID 32
697 - 678 SID 31
400 - 420 · SID 30

RESULT 5
AAS32785
31-39
ID AAS32785 standard; DNA; 11853 BP.
XX
AC AAS32785;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human genomic DNA for novel endocrine antigen, SEQ ID No 739.
XX
KW Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX

OS Homo sapiens.
XX
PN WO200155319-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01335.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457726/49.
XX
PT Isolated polypeptide for treating, preventing and prognosing disorders
PT related to the endocrine system including endocrine disorders,
PT reproductive disorders, and gastrointestinal disorders and also for
PT testing and detection e.g. diagnosis -
XX
PS Disclosure; SEQ ID No 739; 558pp; English.
XX
CC The invention relates to cDNAs encoding novel human endocrine
CC antigens or a fragment having biological activity, a domain, an epitope,
CC full length protein, variant, allelic variant or a species homologue of
CC the cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition when administered
CC (e.g. by gene therapy or antisense-therapy). Identifying mutations in
CC the genes coding for the antigens is useful for diagnosing a pathological
CC condition or a susceptibility to a pathological condition. The DNAs,
CC antigens and antibodies raised against the antigens useful for treating,
CC preventing and/ or prognosing disorders related to the endocrine system
CC or hormone imbalance or reproductive disorders, cancers of endocrine
CC tissues, disorders of the pancreas (e.g. diabetes mellitus), the adrenal
CC glands (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the

```
CC   hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC   of diseases and disorders are given in the specification. The present
CC   sequence is genomic DNA fragment form a gene encoding an endocrine
CC   antigen of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 11853 BP; 3002 A; 3353 C; 2845 G; 2653 T; 0 other;


  Query Match             100.0%;  Score 21;  DB 22;  Length 11853;
  Best Local Similarity   100.0%;  Pred. No. 1;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gtaaaaaatcatgagccctgc 21
           |||||||||||||||||||||
Db    1370 gtaaaaaatcatgagccctgc 1390


RESULT   6
AAD07609/c
ID   AAD07609 standard; cDNA; 546 BP.
XX
AC   AAD07609;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 8 cDNA clone HATDM46, SEQ ID NO:49.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             131..337
FT                   /*tag=  a
FT                   /product= "Human secreted protein precursor"
FT   sig_peptide     131..184
FT                   /*tag=  b
FT   mat_peptide     185..334
FT                   /*tag=  c
FT                   /product= "Mature human secreted protein"
XX
PN   WO200132676-A1.
XX
```

2009 − 1990 = SID 38
has 37
36
35
34
33
32
31

39 only

PD 10-MAY-2001.
XX
PF 25-OCT-2000; 2000WO-US29365.
XX
PR 29-OCT-1999; 99US-0162237.
PR 21-JUL-2000; 2000US-0219666.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Shi Y, Olsen HS, Soppet DR;
XX
DR WPI; 2001-328773/34.
DR P-PSDB; AAE03090.
XX
PT Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease,
PT diabetes mellitus and multiple sclerosis -
XX
PS Claim 1; Page 434; 546pp; English.
XX
CC AAD07571-AAD07645 represent cDNAs corresponding to 25 human secreted
CC protein genes, and AAE03052-AAE03126 represent the proteins they encode.
CC AAE03127-AAE03150 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 25 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 546 BP; 131 A; 120 C; 118 G; 173 T; 4 other;

Query Match 80.0%; Score 16.8; DB 22; Length 546;
Best Local Similarity 90.0%; Pred. No. 84;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       2 taaaaaatcatgagccctgc 21
           |||| |||||||||| ||||
Db     316 TAAATAATCATGAGCTCTGC 297
```